Testing the Limits of Face Recognition:

Identification from Photographs in Travel Documents

and Dynamic Aspects of Emotion Recognition

presented to the Faculty of Arts
of
the University of Zurich
for the degree of Doctor of Philosophy

by
Sarah Chiller-Glaus
of
Zurich

accepted in the fall semester 2008
on the recommendation of
Prof. Dr. Wolfgang Marx and Prof. Dr. Fred Mast

Zentralstelle der Studentenschaft
Zürich 2009

CONTENT

Part III: Emotion Recognition

Summary and Outline

The following work aims at investigating the recognition of facial identity and expression. It is structured in three parts:

In **Part I**, an introduction to the basic concepts of face recognition is given, including an overview of the state of the art literature on both psychophysical and neurological research on face recognition. The concepts of configural and part-based information contained in faces, the effect of inversion, differences between identity and expression recognition, and dynamic information are discussed, as well as regions in the brain associated with face recognition, and three prominent neurological models of face recognition. It is the aim of this part to set the theoretical groundwork for the experiments described in Part II and III.

Part II, which is the main part of this work, focuses on an applied aspect of face recognition, i.e. the field of identity verification at border control. It contains three chapters. In the first chapter, a series of experiments show that verifying the identity of a person by means of a document photograph is highly prone to errors: *Experiment 1* reveals that time pressure at border control can have a detrimental influence on identification performance. *Experiment 2* addresses the issue of the size of the document photograph and whether it is sufficiently large for reliable identification. In *Experiment 3*, a variety of security personnel are tested to investigate the role of expertise in identity verification. The experiment includes both Asian and Caucasian faces to analyze whether the race of the document holder influences identification. Also, the possible effects of inversion of the photographs are assessed. The experiment reveals that despite several years of experience, none of the experts perform better than untrained laymen. Implications on the maintenance of security at border control are discussed.

The second chapter then focuses on ways of how to ameliorate the situation. *Experiment 4* reveals that arts students can identify faces better after a seven-week's course in portrait painting. This indicates that identification from photographs is indeed trainable. With *Experiment 5* therefore a training system is developed. Identification turns out to be better after training, but the effect is only small.

The last chapter of Part II addresses methodological questions which arise from Experiments 1-5. In *Experiment 6* a simultaneous matching paradigm is compared to old-new recognition, and in *Experiment 7* the use of siblings as stimulus material is critically assessed.

Finally, **Part III** returns to basic research and focuses on the recognition of facial expression, in particular the role of dynamic information. Using a composite-face paradigm, *Experiment 8* shows that motion does not separately enhance configural or part-based information, but seems to have a quality of its own.

The two approaches – basic and applied research, recognition of both expression and emotion – cover a wide range of aspect relevant for face recognition and are thus hoped to significantly contribute to the current research on one of the most relevant objects known to us – the human face.

Introduction

Recognizing people's identity and their emotional state is a basic and important skill in social interaction which we perform with great accuracy and consistency.

Indeed, out abilities in face recognition are astonishing: We are able to identify a familiar face out of a crowd of other people with highest accuracy. We are able to recognize a person under a tremendous variety of chances, such as different hairstyle, change of glasses or jewellery, of headdress or clothes, or even after the passing of many years. Also subtle changes, such as the position of the head, changes in expression, in lightning, or faces in motion do not challenge our abilities. Regarding the fact that all faces share the same basic arrangement, eyes above mouth and nose in the middle, such skills are truly marvelous. There is by far no other class of objects which we are able to recognize so accurately and reliably like the human face.

Most of the time, however, these skills go unnoticed. It is only in situations when correct recognition for once does not work that we become aware of it: How unsettling, when we are greeted by a person and cannot immediately place her! How embarrassing when we approach someone with a large smile and realize at closer look that we are mistaken in the person! A very striking report is given by a writer who suffers from prosopagnosia, a disorder of face recognition where faces cannot be perceived correctly:

> *"…Was mich am meisten verwirrt: Ich sehe doch Gesichter – Augen, Nase und Mund –, kann sie sogar beschreiben. Wenn ich [mein Gegenüber, scg] anschaue, erkenne ich an seinen schmalen Augen sofort seine asiatische Abstammung, er hat für einen 67-Jährigen ein jugendliches Gesicht, eine breite Nase und schmale Lippen. Wie kann es sein, dass ich ihn auf der Strasse nicht mehr erkennen würde?" (Schneider, 2008).*

Such narrations impressively show the importance of our intact ability to recognize faces. And not just the identity of a person, but also somebody's emotional state: We are able to read in other people the most subtle nuances of emotion and interpret them in ways meaningful to us: Have we angered the person? Is he happy? Does someone agree with what I said, do I look at friend or fiend? Facial emotions, and with them our ability to correctly and reliably read them, are highly crucial for social interaction. Someone's face is our most authentic mirror to the other person's emotional state.

It is surely this outstanding position that faces take up in our lives that has rendered them a prominent topic for scientific research. Probably the most urgent focus has been the question of how we perceive faces; of what is it that makes them so special and why we are so good at recognizing them. Both behavioral and neurological studies have addressed this issue thoroughly. The current position holds that the reason for our exceptional abilities is the fact that we process a face's configurations, i.e. the spatial distances between the separate parts (eyes, nose, mouth, etc.), while with other objects we focus mainly on the separate parts themselves. This theoretical knowledge of our special way to process faces has an influence on many other areas of research, for example the question why people from other races seem to look so similar to each other that we have difficulties in holding them apart; or the explanation of certain illusion and effects like the Thatcher illusion, the Composite Face Effect, or the Face Inversion Effect. This work attempts to give an overview over the relevant research in face recognition and explain its basic concepts (Part I). An area which has not been researched in detail yet is what we make of the fact that in most experiments photographs of faces are used, while in reality we are confronted with living, and moving, faces. How does motion affect this special way of ours to process faces? Part III addresses this question.

A relatively new field of face recognition which has not yet been well-researched is the identification of a face from a photograph. It is a common source of exhilaration to look at old class mates or family photos and try to make out in the group the person who today looks very different. Contrary to our exceptionally high abilities in recognizing a living person as described above, however, the identification from a photograph is not always so easy. Who has not made the experience of looking at an old photo of a friend and unbelievingly exclaiming: What, this is you? And who has not had doubts whether the train conductor checking the ticket was at all able to ascertain the ownership of the ticket by a small, black-and white, and maybe out of date photograph? These situations clearly indicate that despite all our skill and aptitude, there are limits to face recognition. To address the question of where these limits lie and whether they can be abolished is the main research question of this work. It shall be discussed in Part II.

Part I

An Introduction in Face Recognition

1. Psychophysical Perspective of Face Processing

1.1. Components and Configurations

A common classification of the information contained in faces is the distinction between *component* information, relating to separable local elements such as eyes, mouth, or nose (Carey & Diamond, 1977; Sergent, 1984), and *configural* information, referring to the spatial relations of these elements (Bruce, 1988; for reviews see Schwaninger, Carbon, & Leder, 2003; Schwaninger, Wallraven, Cunningham, & Chiller-Glaus, 2006). Both terms have many synonyms in literature. Component information is also, among other expressions, referred to as “part-based” (e.g. Calder, Young, Keane, & Dean, 2000), “featural” (e.g. Young, Hellawell, & Hay, 1987), “componential”, or “piecemeal” information (see Schwaninger et al., 2003), configural as “second order spatial relations” (e.g. Diamond & Carey, 1986), “holistic” (e.g. Bartlett & Searcy, 1993), “configurational” (e.g. Collishaw & Hole, 2000) or “relational information” (e.g. Leder & Bruce, 1998), which do not always precisely refer to the same concept. Especially the use of the term “holistic” for “configural” seems problematic, since it also stands for other concepts than configural information[1]. It is therefore the term “configural” that is used in this work, and it refers to the spatial relations between a face’s components. The term “part-based” is used for the processing of the components themselves.

Several hypotheses have been proposed to explain the mechanisms used in adult face processing, although general consensus holds that for the recognition of faces, configural information is of special importance. Two prominent hypotheses shall be discussed in the following.

1.2. Holistic versus Part-Based Processing

Two main hypotheses have been proposed to explain the recognition of identity in faces: The holistic and the component-configural hypothesis. According to the *holistic hypothesis*, upright faces are stored as unparsed perceptual wholes in which components are not explicitly represented (Farah, Tanaka, & Drain, 1995; Farah, Wilson, Drain, & Tanaka, 1998; Tanaka & Farah, 1993). Note that the term “holistic” here refers not to configural information alone as

[1] According to Tanaka and Farah (1993), „holistic“ refers to the perception of a face as unparsed whole in which no components (likely also no configurations) are explicitly represented. Tanaka and Sengco (1997) and

suggested by Bartlett and Searcy (1993), but to the face as a whole. The main empirical evidence in favor of the holistic hypothesis is based on a paradigm by Tanaka and Farah (1993). They argued that if face recognition relies on parsed representations, then single parts of a face, such as nose, mouth or eyes, should be easily recognized even if presented in isolation. However, if faces are represented as unparsed perceptual wholes (i.e. holistically), then recognition of the same isolated parts should be more difficult. In their experiments, participants were shown a previously learned face together with a slightly different version in which one single part (e.g., nose or mouth) had been replaced. The task was to judge which face has been shown in the learning phase. The experiment was conducted in both a whole face condition and in an isolated parts condition without facial context. In the isolated condition, face parts proved to be more difficult to recognize than in the whole face condition. However, when participants were trained to recognize other objects such as houses no advantage of context was found. Tanaka and Farah concluded that face recognition relies mainly on holistic representations, in contrast to the recognition of objects.

An alternative hypothesis is the *component-configural hypothesis*. It assumes that face recognition relies on explicit representations of both component and configural information (e.g. Sergent, 1984; Searcy & Bartlett, 1996; Leder & Bruce, 2000; Williams, Moss, & Bradshaw, 2004; for recent overviews see Schwaninger et al., 2006). Of these two types of information, it is the configural which plays a pivotal role in face recognition, while the components are less critical, but nevertheless processed explicitly.

Figure 1.1: Composite face effect

Aligned (left) and misaligned (right) composites of Marilyn Monroe (top) and Margaret Thatcher (bottom). The identity of both facial halves is more difficult to recognize when aligned than when misaligned.

Schwaninger, Carbon, and Leder (2003), among others, describe “holistic” as integration of both configural and part-based information into a single face representation.

The so called composite face effect (Young, Hellawell, & Hay, 1987) is an impressive demonstration of this circumstance[2] (see Figure 1.1): When combining the top half of one face with the bottom half of another face in alignment, recognition is significantly impaired in respect to misaligned halves. This fact is generally explained by the fusing of the aligned halves to one single identity, while in the misaligned version the configural information is disrupted. So, the configural information seems to be the main source upon which we draw when processing a face. At the same time, the components are not completely negligible, since the identity of the two misaligned halves is still recognizable despite lack of configural information.

Many studies regarding the component-configural hypothesis changed component information by replacing components (e.g. eyes were replaced with the eyes of another person). Configural changes were induced by altering the distance between components (e.g. larger or smaller inter-eye distance). However, one possible caveat is that these types of manipulations often change the holistic aspects of the face too, and they are difficult to carry out selectively. For example, replacing the nose (component change) might change the distance between the contours of the nose and the mouth, which induces a configural change (Leder & Bruce, 1998; 2000). Moving the eyes apart (configural change) can lead to an increase in size of the bridge of the nose, which is a component change (see Leder, Candrian, Huber, & Bruce, 2001). Such problems can be avoided by using scrambling and blurring procedures to reduce configural and component information independently (e.g., Boutet, Colin, & Faubert, 2003; Collishaw & Hole, 2000; Davidoff & Donnelly, 1990; Sergent, 1985). Using such techniques, Schwaninger, Lobmaier, and Collishaw (2002) could show that previously learned intact faces could be recognized even when they were blurred (elimination of components), or when they were scrambled into constituent parts (elimination of configurations), indicating that both configurations and components played a role in face recognition. This result challenges the assumption of purely holistic processing according to Farah et al. (1995) and suggests that components are encoded and stored explicitly.

---

[2] It is interesting to note that while some authors take the composite face effect as evidence for holistic processing (see White, 2000), Young et al. (1987) themselves explicitly state that „configurational and featural information are […] both likely to contribute to normal face recognition“ (p. 758).

1.3. Inversion of Faces

One characteristic attribute of face recognition is that the processing of faces is highly orientation-dependent: Yin (1969) was the first to describe that upside-down faces are disproportionately more difficult to recognize than other inverted objects, a finding which is generally referred to as *face inversion effect* (FIE). Subsequent replications of Yin's study have confirmed this result (e.g., Ellis, 1975; Goldstein & Chance, 1981; Leder & Bruce, 1998; Rhodes, Brake, & Atkinson, 1993; Scapinello & Yarmey, 1970; Searcy & Bartlett, 1996; Sergent, 1984; Tanaka & Farah, 1991; see Schwaninger et al., 2003, or Valentine, 1988, for a review). The mechanisms responsible for the face inversion effect can well be explained by the already mentioned composite face paradigm: As described above, aligned upright face composites result in the perception of a new identity, presumably by the processing of configural information which is available in aligned, but not in misaligned composites. In aligned composites, it is therefore more difficult to perceive the identity of each half than in misaligned composites. In inverted composites, however, no such effect can be seen, i.e. no difference is found between aligned and misaligned composites (Young et al., 1987). This finding indicates that in inverted faces, configural information is no longer accessible, while the processing of components seems relatively unaffected: The missing difference between aligned and misaligned composites suggests that both objects are processed in similar ways, namely by their separate parts, while configural information – which if processed should result in the perception of a new identity in aligned composites and therefore reduce the identification of either half – is not available.

Another phenomenon to explain the mechanisms responsible for the face inversion effect is the so called *Thatcher illusion:* When rotating the eyes and mouth in an otherwise unchanged face, a grotesque expression appears. This expression, however, is much less severe when the face is turned upside down than when upright (see Figure 1.2). This illusion was discovered by Thompson (1980) using the photograph of the English then-prime minister Margareth Thatcher, hence the name "Thatcher Illusion". Again, configural information seems to play a pivotal role: In the inverted version, it is no longer accessible, and a relatively normal expression appears. The face has to be processed on the basis of its features, which are oriented normally in the inverted version and therefore seem unsuspicious. Lacking configural information, our system does not integrate the features in the surrounding background, which would allow realizing that the object was highly unusual. It is therefore only in the upright version that we become aware of the grotesque expression.

Figure 1.2: Thatcher Illusion

In the left image, eyes and mouth are turned 180° in respect to the surrounding facial background. When seen upside down, the expression in both faces is relatively normal, although the left one is somewhat unusual. Only when viewing the images right side up the grotesque expression of the left image becomes apparent.

In sum, inverted faces are more difficult to recognize than upright faces, and this seems to be mainly due to the disruption of configural information. Schwaninger and Mast (2005) explain it as follows: since in inverted faces configural information is not easily accessible anymore, face recognition has to draw on the separate facial features and rotate them piece by piece. Due to the complexity of this task, however, at large rotation angles (such as 180°, i.e. inversion) this mechanism is overtaxed and performance is reduced – a concept which goes back to Rock (1973, 1974, 1988).

1.4. Identity and Expression

The vast majority of studies concern the recognition of identity of faces. Only comparably few studies have research the recognition of emotion. Calder et al. (2000) explain this circumstance by the fact that until the 1980s facial expressions were researched mainly in the domain of social and not cognitive psychology. Only in the last two decades did the research of emotion recognition gain momentum in the field of cognitive psychology too.

The above described concepts in face recognition – components and configurations, inversion – apply to both identity and expression recognition. Configural information is not only central in the recognition of facial identity, as has been discussed so far, but also in the processing of emotions. Although many researchers have pointed at the functional differences between the recognition of identity and of facial expression (e.g. Bruce & Young, 1986; O'Toole, Roark,

& Abdi, 2002, see Chapter 2), in terms of configurations and components, the same mechanisms seem to apply to expression recognition as to identity recognition. McKelvie (1995) could show that emotions which were easily recognized in upright orientation could not be named when inverted. Inversion being known to disrupt configural processing (Yin, 1969), the results of McKelvie thus lead to the assumption that also in emotion processing configural information is of central importance. Similar results were attained by Prkachin (2003). Furthermore, Young et al.'s composite paradigm, originally designed for facial identity, proved to be also applicable to expression recognition (Calder et al., 2000).

But besides the importance of configural information, the main focus of emotion recognition research lies on the separate features. It is a central assumption in emotion recognition that different facial areas are important for the recognition of different emotions (Bassili, 1979; Cunningham, Kleiner, Wallraven, & Bülthoff, 2005; Hanawalt, 1944; Nummenmaa, 1964; Plutchik, 1962). For example, Bassili (1979) used point-light faces to show that the upper portions of the face are important for some expressions, while the lower portions of the face are important for other expressions. Facial features also play differentiated roles in other aspects of facial expression processing, such as the perception of sincerity. For example, according to Ekman and Friesen (1982), a true smile of enjoyment, which Ekman refers to as a Duchenne smile, has a characteristic mouth shape as well as specific wrinkles around the eyes. Faked expressions of enjoyment, in contrast, contain only the mouth information. Furthermore, Ekman and Friesen (1982) have shown that deceptive expressions of enjoyment appear to have different temporal characteristics than spontaneous ones.

Modelling Facial Expression Recognition

Given the general pre-occupation with the role of part-based information in the recognition of facial expressions, it is not surprising that the vast majority of descriptive systems and models of facial expressions are explicitly parts-based (Ekman & Friesen, 1978; Elison & Massaro, 1997; Essa & Pentland, 1994; Frijda & Philipszoon, 1963; Izard, 1979; Leventhal & Sharp, 1965; Tronick Als, & Brazelton, 1980). Perhaps the most widely used methods for parameterizing the high-dimensional space of facial expressions is the facial action coding system (or FACS, Ekman & Friesen, 1978), which segments the visible effects of facial muscle activity and rigid head motion into "action units". Combinations of these action units can then be used to describe different expressions. It is important to note that FACS was designed as a system for describing the elements of photographs of facial expressions. It is not

a model of facial expression processing and makes no claims about which elements go together to produce different expressions (Sayette, Cohn, Wertz, Perrott, & Dominic, 2001).
Elison and Massaro (1997) proposed a parts-based model of perception (the fuzzy logical model of perception, or FLMP) in which the features are independently processed and subsequently integrated. The model makes specific claims about how the part-based information is processed and integrated, and thus makes clear predictions about the perception and categorization of facial expressions. In one study, Elison and Massaro used computer graphics animation techniques to produce static facial expressions where either (a) the mouth shape was parametrically varied; (b) the eyebrow shape was parametrically varied, or (c) both were independently parametrically varied. The faces were shown to a number of observers, who were asked if the expression in the photographs was happy or angry. Elison and Massaro found that both features (eyebrow position and mouth position) affected the participants' judgments, and that the influence of one feature was more prominent when the other feature was neutral or ambiguous. Moreover, the FLMP captured patterns in the data better than either holistic models or a straight forward additive model based on recognition rates of the individual features. Elison and Massaro consequently claimed that the perceptual system must be using part-based information in the recognition process and can not be employing a purely holistic approach. These results are consistent with the aforementioned findings on identity recognition.

Apart from the part-based models described above, there are at least two models that integrate holistic information (Izard, Dougherty, & Hembree, 1983; White, 2000). White (2000) proposed a "hybrid model", according to which expression recognition is part-based on one hand and holistic in the sense of undecomposed wholes on the other hand.

Taken together, the recognition of facial expressions seems to be based on the same principles as identity recognition: It is most likely a combination of both part-based and configural information, rather than purely holistic processing, which guarantee our reliable and fine-tuned recognizing of facial expressions.

1.5. Dynamic Information

In reality, other than in most experiments on face recognition using mere photographs, faces are not static objects, but are constantly in motion. Compared to the wealth of research on emotion recognition, however, only a relatively small number of studies have addressed the role of dynamic information (Ambadar, Schooler, & Cohn, 2005), i.e. whether faces are recognized better when seen in motion rather than statically. One traditional way of describing motion is to separate it into rigid and non-rigid motions (see, e.g., Gibson, 1957, 1966; Roack, Barret, Spence, Abdi, & O'Toole, 2003). Rigid face motion generally refers to the rotations and translations of the entire head (such as occurs when nodding the head). Non-rigid face motion, in contrast, generally refers to motion of the face itself, which consists mostly of non-linear surface deformations (e.g., lip motion, eye brow motion). Most naturally occurring face-related motion contains both rigid and non-rigid motion. Motion could contribute to face recognition by several mechanisms, e.g. by supplemental information due to an increased number of views available, by building up a 3D-representation, or by a quality of its own inherent to dynamic information (e.g., Lander & Bruce, 2000, for the recognition of facial identity; O'Toole et al., 2002, for face recognition in general).

Does Motion Facilitate Face Recognition?

For both the recognition of identity and expression, however, research up to date has not established a clear answer weather motion facilitates recognition or not (O'Toole et al., 2002). Evidence for a beneficiary effect of motion in *identity* recognition is given for example by Knappmeyer, Thornton and Bülthoff, (2003) or Lander and Bruce (2004, 2000). Also, under suboptimal viewing conditions such as poor illumination or long distance, dynamic information proved to be helpful (Lander, Christie, & Bruce, 1999). There are, however, converse findings: Bruce, Henderson, Greenwood, Hancock, Burton, and Miller (1999) demonstrated difficulties in matching unfamiliar target faces on video against arrays of photographs, where accuracy proved to be poor in the static condition even when viewpoint and facial expression were standardized, and did not improve when the target face was shown in motion. Christie and Bruce (1998) confirmed the lack of improvement.

Like with identity recognition, there are ambiguous findings on the role of dynamic information also for the recognition of *expressions*: In a direct examination of the role of motion, Bassili (1978, 1979) used Johannson point-light faces as stimuli (see Johannson, 1973, for more on point-light stimuli). He could show that facial expression undergoing dynamic change were perceived correctly even when part-based information was eliminated, while this was not the case for emotions under static conditions. Kamachi, Bruce, Mukaida, Gyoba, Yoshikawa, and Akamatsu (2001) manipulated the velocity in which a neutral face turned into an emotional one. They found that happiness and surprise were better recognized from fast sequences, sadness better from slow sequences, concluding that – depending on the velocity of change – motion assisted emotion recognition, due to the representations of emotions encoding information about both dynamic and static properties. The overall performance was nevertheless slightly poorer for dynamic images than for static images, which was explained by the fact that in the static condition, 100% of the target emotion was present for the full display duration, while in the moving condition an average of only 50% of the target emotion was presented for the entire sequence.

Possible Modes of Effect of Dynamic Information

Apart from the question whether dynamic information facilitates face recognition, there still is a debate over the exact mechanisms by which dynamic information should contribute to face recognition – in other words: *how* it helps. Lander et al. (1999), Lander and Bruce (2000), and Pike et al. (1997) performed a series of experiments to ensure that the apparent advantage of moving faces over static faces is due to information that is solely available over time (i.e., dynamic information). One might, for example, describe dynamic sequences as a series of static snapshots. Under such a description, the advantage of dynamic stimuli would not lie with dynamic information, but with the fact that a video sequence has more static information (i.e., it has supplemental information provided by the different views of the face). To test this hypothesis, Lander et al. (1999) presented their participants three versions of a 9-frame video sequence of famous faces: once the original video sequence, once the nine frames as static images in ordered array, and once the same static images in jumbled array. The faces were recognized better in the video condition than in the either of the two static conditions, and performance in the two static conditions did not differ from one another. Thus, it seems that video sequences are not better simply because they contain more snapshots. To test whether the advantage is due to motion in general, or due to some specific type of motion, Lander & Bruce (2000) and Pike et al. (1997) presented a video where the images were in a random order. Note that such sequences have motion information, but this motion is random (and does not occur in nature). It was found that identity was more accurately recognized in the normal sequences that in the random sequences, implying that not just the presence of motion that is important, but the specific, naturally occurring motion that provides the advantage. Further, it was found that reversing the direction of motion (by playing the sequence backwards) decreases recognition performance, suggesting that the temporal direction of the motion trajectories is important too (Lander & Bruce, 2000). By changing the speed of a motion sequence (e.g., by playing parts or all of a video sequence too fast or slow), the researchers showed that the specific tempo and rhythm of motion is important for face recognition

(Lander & Bruce, 2000). Finally, a study conducted by Ambadar et al. (2005) revealed motion enhanced the perception of change in faces, which should explain the better recognition of faces in motion than of static images.

In this chapter, the basic concepts of face recognition were introduced. It was shown that the information contained in faces is commonly classified in part-based information and configural information, the latter being most central for reliable recognition. Inversion disrupts configural information and thus reduces recognition drastically (face inversion effect). Some form of facial information seems to be available only over time, as the majority of studies on dynamic information revealed. And finally, these concepts seem to apply to both the recognition of a face's identity and its emotional state.

2. Physiological Perspective

2.1. Face-Selective Areas – Evidence from Neuroscience

At least since the discovery of the face inversion effect (Yin, 1969) the question has been discussed whether a specific area for the processing of faces exists in the human brain. Neuropsychological evidence for specialization has been derived from prosopagnosia, a deficit in face identification following inferior occipitotemporal lesions (e.g. Damasio, Damasio, & van Hoesen, 1982; for a review see De Renzi, 1997). There have been a few reports of prosopagnostic patients in which object recognition seemed to have remained intact (e.g. Bentin, Deouell, & Soroker, 1999; Farah et al., 1995; McNeil & Warrington, 1993). Prosopagnosia has been regarded as a face-specific deficit which does not necessarily reflect a general disorder in exemplar recognition (e.g. Henke, Schweinberger, Grigo, Klos, & Sommer, 1998). Consistent with this view, patients have been reported that suffered from associative object agnosia, while their face identification remained unaffected (e.g. Moscovitch, Winocur, & Berhmann, 1997). Such a double dissociation between face and object recognition would imply that the two abilities are functionally distinct and anatomically separable. However, based on methodological concerns, some authors have doubted whether face recognition can really be dissociated from object recognition based on current literature on prosopagnosia (e.g. Gauthier, Behrmann, & Tarr, 1999a, see also Davidoff, & Landis, 1990).

Evidence for the uniqueness of face processing has also been derived from ERP and MEG studies. A response component called the N170 (or M170 in MEG) occurring around 170 ms after stimulus onset, is usually twice as large for face stimuli when compared to other control stimuli such as hands, houses or animals (e.g. Bentin et al., 1996; Liu, Harris, & Kanwisher, 2002). However, the debate on whether such activation is unique for faces or whether it represents effects of expertise that are not specific to face processing is still ongoing (for recent discussions see for example Rossion, Curran, & Gauthier, 2002; Xu, Liu, & Kanwisher, 2005).

FFA, STS and OFA

In functional brain imaging, several areas have been identified to be of special importance for the processing of faces (see Haxby, Hoffman, & Gobbini, 2000 for a review). These involve a region in the in the lateral fusiform gyrus, the superior temporal sulcus (STS), and the "occipital face area" (OFA, Gauthier, Tarr, Moylan, Skundlarski, Gore, & Anderson, 2000a).

All areas have been identified bilaterally, albeit with a somewhat stronger activation in the right hemisphere. The face-selective area in the fusiform gyrus has been referred to as the “fusiform face area” (FFA) by Kanwisher, McDermott, and Chun (1997). While FFA activation has been related to facial identity, the STS in humans reacts particularly to changing aspects of faces with social value, such as expression, direction of gaze and lip movement (e.g. Hoffman & Haxby, 2000; Puce, Allison, Bentin, Gore, McCarthy, 1998). In a fMRI study using adaptation (reduction of brain activity due to repetitive stimulus presentation), Andrews and Ewbank (2004) investigated differences between the FFA and the STS in face processing. Activity in the FFA was reduced over time by stimuli of the same identity. Adaptation was dependent on viewpoint but not on size changes. Adaptation to identity was not found in STS but an increased response when the same face was shown with a different expression or from different viewpoints. These results suggest a relatively size-invariant neural representation in FFA for recognition of facial identity, and a separate face-selective region in STS involved in processing changeable aspects of a face such as facial expression. OFA and inferior occipital gyrus seem to be associated with early structural encoding processes; they are primarily sensitive to sensory attributes of faces (Rotshtein, Henson, Treves, Driver, & Donlan, 2005). Rossion et al. (2003) obtained results in an fMRI study suggesting that OFA and FFA might be functionally associated: PS, a patient suffering from severe prosopagnosia due to lesions in the left middle fusiform gyrus and the right inferial occipital cortex, performed poorly in a face matching task despite normal activation of the intact right FFA. Rossion et al. thus conclude that the FFA alone does not represent a fully functional module for face perception, but that for normal face processing intact OFA and FFA in the right hemisphere with their re-entrant integration are necessary. Yovel and Kanwisher (2005), however, came to a different conclusion. They correlated the behavioral performance in a face matching task of upright and inverted faces with the neuronal responses to upright and inverted faces in the three regions FFA, STS and OFA. It was found that only the FFA showed a difference in activity between upright and inverted faces. This can be interpreted as functional dissociation between FFA and the other cortical regions involved in face processing. The authors also conclude that the FFA appears to be the main neurological source for the behavioral face inversion effect originally reported by Yin (1969). The latter however, is not exclusive to faces. In a behavioral study, Diamond and Carey (1986) found comparable inversion effects for faces and side views of dogs when dog experts were tested. Subsequent behavioral and imaging studies using recognition experiments with trained experts and artificial objects (“Greebles”), as well as bird and car experts with bird and car images provided further evidence in favor of a process-specific rather than a domain-specific interpretation (Gauthier et al, 1999b, 2000b). According to their view (“expertise hypothesis”), FFA activity is related to the identification of different classes of visual stimuli if they share the same basic configuration and if substantial visual expertise is given. The question on whether FFA activity is domain or process-specific is being debated since several years now. It is beyond the scope of this chapter to review this ongoing debate but for an update on the current status see for example, Chan, Peelen, Dodds, & Kanwisher (2005), Xu (2006), Bukach, Gauthier, & Tarr (2006), Kanwisher and Yovel (2006).

Nevertheless, it should be noted that activation in face-selective regions of the fusiform area is not exclusive to faces. Significant responses to other categories of objects have been found in normal subjects, for example for chairs, houses, and tools (Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Ishai, Ungerleider, Martin, & Haxby, 2000; Haxby, Gobbini, Furey, Ishai, Schouten, & Pietrini, 2001). Moreover, it has also been shown that face-selective regions in the fusiform area can be modulated by attention, emotion and visual imagery, in addition to modulation by expertise as mentioned above (e.g. Ishai, Haxby, & Ungerleider, 2002; O’Craven, Downing, & Kanwisher, 1999; Vuilleumier, Armony, Driver, & Donlan, 2001).

2.2. Cognitive Neuroscience Models of Face Processing

In recent years, substantial progress has been made regarding models on how different brain areas interact in processing information contained in faces. Three main accounts are summarized in the following section.

1. The model by Bruce and Young (1986) is one of the most influential accounts in the psychological face processing literature. This framework proposes parallel routes for recognizing facial identity, facial expression and speech-related movements of the mouth. It is a rather functional account since Bruce and Young did not provide specifics regarding the neural implementation of their model. The recent physiological framework proposed by Haxby et al. (2000) is consistent with the general conception proposed by Bruce and Young. According to Haxby et al.’s model, the visual system is hierarchically structured into a core and an extended system. The core system comprises three bilateral regions in occipitotemporal visual extrastriate cortex: Inferior occipital gyrus, lateral fusiform gyrus, and STS. Their function is the visual analysis of faces. Early perception of facial features and early structural encoding processes are mediated by processing in inferior occipital gyrus. The

lateral fusiform gyrus processes invariant aspects of faces, as the basis for the perception of unique identity. Changeable properties such as such as eye gaze, expression, and lip movement are processed by STS. The representations of changeable and invariant aspects of faces are proposed to be independent of one another, consistent with the model by Bruce and Young. The extended system contains several regions involved in other cognitive functions such as spatially directed attention (intraparietal sulcus), prelexical speech perception (auditory cortex), emotion (amygdala, insula, limbic system), and personal identity, name, and biographical information (anterior temporal region).

2. The model of Haxby et al. has been taken as a framework for the extension by O'Toole, Roark, and Abdi (2002). By taking into account the importance of dynamic information in social communication, they further explain the processing of facial motion. In their system, dynamic information is processed by the dorsal stream of face recognition, static information by the ventral stream. Two different types of information are contained in facial motion: social communication signals such as gaze, expression, and lip movements, which are forwarded to the STS via the middle temporal (MT) area; and person-specific motion ("dynamic facial signatures"). O'Toole et al. suggest that the later type of information is also processed by the STS, representing an additional route for familiar face recognition. This model is in accordance with the supplemental information hypothesis which claims that facial motion constitutes additional information to static information. According to O'Toole et al., structure-from-motion may also support face recognition through communication between the ventral and the dorsal streams. For instance, the structural representation in FFA could be enhanced by input from the middle temporal area. Thus, the model integrates also the representation enhancement hypothesis.

3. In a detailed review of psychological and neural mechanisms, Adolphs (2002) provides a description of the processing of emotional facial expressions as a function of time. The initial stage provides automatic fast perceptual processing of highly salient stimuli (e.g. facial expressions of anger and fear). This involves the superior colliculus and pulvinar, as well as activation of the amygdala. Cortical structures activated in this stage are V1 and V2, and other early visual cortices that receive input from the lateral geniculate nucleus of the thalamus. Then, a more detailed structural representation of the face is constructed until about 170 ms. This processing stage involves the fusiform gyrus and the superior temporal gyrus, which is consistent with Haxby et al.'s core system. Dynamic information in the stimulus would engage the middle temporal area, middle superior temporal area, and posterior parietal visual cortices. Recognition modules for detailed perception and emotional reaction involve Haxby

et al.’s extended system. After 300 ms conceptual knowledge of the emotion signaled by the face is based on late processing in the fusiform and superior temporal gyri, orbitofrontal and somatosensory corcices, as well as activation of the insula.

Dissociation between Identity and Expression Recognition

It is a common assumption of all these models that identity and expression recognition are processed separately. Neuropsychological evidence suggests a double dissociation, some patients show impairment in identity recognition but normal emotion recognition, and other patients show intact identity recognition but impaired emotion recognition (for reviews see Adolphs, 2002; Damasio et al., 1982; Damasio, Tranel, & Damasio, 1990; Wachholz, 1996). In their study, Winston, Henson, Fine-Goulden, & Donlan (2004) revealed dissociable neural representations of identity and expression using an fMRI adaptation paradigm. They found evidence for identity processing in fusiform cortex and posterior superior temporal sulcus (STS). Coding of emotional expression was related to a more anterior region of STS. Bobes, Martin, Olivares, & Valdés-Sosa (2000) showed that emotion matching resulted in different ERP scalp topography than identity matching. In another ERP study, Eimer and Holmes (2002) investigated possible differences in the processing of neutral versus fearful facial stimuli. They found that the N170, which is related to structural encoding of the face in processing identity, did occur in both the neutral and the fearful condition. This indicates that structural encoding is not affected by the presence of emotional information and is also consistent with independent processing of facial expression and identity. However, results from other studies challenge the assumption of completely independent systems. DeGelder, Frissen, Barton, & Hadjikhani (2003) found that subjects suffering from prosopagnosia performed much better when faces showed emotions than when they depicted a neutral expression. With normal subjects, the opposite was the case. DeGelder et al. assume that the areas associated with expression processing (Amygdala, STS, parietal cortex) have a modulatory role in face identification. Their findings challenge the notion that different aspects of faces are processed independently (assumption of dissociation), and only after structural encoding (assumption of hierarchical processing). Calder and Young (2005) share a similar view. They argue that a successful proof of the dissociation of identity and expression would require two types of empirical evidence. First, patients with prosopagnosia but without any impairment in facial expression recognition. Second, intact processing of facial identity and impaired recognition of emotion without impairment of other emotional functions. Based on their review the authors conclude that such clear patterns have not been revealed yet. The reported selective disruption of facial expression recognition would rather reflect an impairment of more general systems than damage (or impaired access) to visual representations of facial expression. The authors do not completely reject the dissociation of identity and expression, but they suggest that the bifurcation takes place at a much later stage than proposed by the model of Haxby et al., namely only after a common representational system. A critical problem of those approaches however, is that they rely on a purely holistic processing strategy of face stimuli, which in light of the previously discussed behavioral evidence seems not plausible.

Components and Configurations

As discussed before, there is a large number of studies in the psychophysical literature that clearly suggests an important role of both component and configural information in face processing. This is supported by neurophysiologic studies. In general, it has been found that cells responsive to facial identity are found in inferior temporal cortex while selectivity to facial expressions, viewing angle and gaze direction can be found in STS (Hasselmo, Rolls, & Baylis, 1989; Perret, Hietanen, Oram, & Benson, 1992). For some neurons, selectivity for particular features of the head and face, e.g. the eyes and mouth, has been revealed (Perret, Rolls, & Caan, 1982; Perret, Mistlin, & Chitty, 1987; Perret et al., 1992). Other groups of cells need the simultaneous presentation of multiple parts of a face, which is consistent with a more holistic type of processing (Perret & Oram, 1993; Wachsmuth, Oram, & Perret, 1994). Yamane, Kaji, & Kawano (1988) have discovered neurons that detect combinations of distances between facial parts, such as the eyes, mouth, eyebrows, and hair, which suggest sensitivity for the spatial relations between facial parts (configural information).

Although they are derived from different physiological studies, the three models by by Haxby, O’Toole et al., and Adolphs share many common features. Nevertheless, it seems that some links to behavioral and physiological studies are not taken up in these models. As discussed above, the concept of component and configural processing seems to be a prominent characteristic of face processing. The models, however, do not make this processing step explicit by specifying at which stage this information is extracted. More research is needed to fill this gap.

Note by the author: The text of Part I is based upon (but not identical with) Schwaninger, A., Wallraven, C., Cunningham, D. W., & Chiller-Glaus, S. D. (2006). Processing of identity and emotion in faces: a psychophysical, psychological and computational perspective. Progress in brain research. 156, 321-343. The author wishes to thank the co-authors of the publication.

Summary

The review of psychophysical studies showed that faces are processed in terms of their components and their spatial relationship (configural information). The recognition of faces is strongly reduced by inversion (face inversion effect; Yin, 1969), this effect being mainly the result of impaired configural processing, while component processing is not so much affected by inversion. Similar mechanisms – processing of components and configurations, impairment by inversion – apply for the recognition of both identity and facial expression. Different facial areas and facial motions are important for the recognition of different emotions. Most of the models on facial expression processing have stressed the importance of part-based information while some models also integrate configural information. Dynamic information has largely an enhancing effect upon face recognition, although there are some studies which could not find such an effect. The benefit of dynamic information seems not to lie in the simple fact that movies contain more images than static frames, but due to a specific quality inherent in dynamic information itself, or due to the enhancement of sensibility to change.

Whether faces are processed in specific face-selective areas in the brain, or whether such areas respond to any object of expertise and not just faces is still subject to debate. But it is clear that there are at least three areas which are pivotal to face recognition (if not to faces alone): The fusiform face area (FFA), the superior temporal sulcus (STS), and the occipital face area (OFA), all identified bilaterally. Separate routes for the processing of facial identity and expression can be assumed, although it is not clear yet at what stage the bifurcation takes place. The model by Bruce and Young (1986) proposes separate parallel routes for recognizing facial identity, facial expression and speech. Recent physiological models proposed by Haxby et al. 2000, O’Toole et al. (2002) and Adolphs (2002) are consistent with this view.

The research discussed so far represent the basic concepts in face recognition which are needed to understand the research conducted in the following two parts of this work. While most of the literature discussed in this introduction involved basic research, Part II focuses on applied aspects of face recognition, in particular the verification of identity by photographs, as is carried out at border control. Part III then closes the circle back to basic research, i.e. the recognition of emotion and their dynamic aspects.

Part II

Document Verification

On a train journey between two towns in Switzerland I witnessed a very comical scene: As I prepared to show my ticket to the conductor coming down the aisle, I was suddenly aware that further down in the carriage a Black passenger had become the center of everybody's attention: Apparently, the conductor was not sure whether the photo on the ticket really depicted the man sitting in front of him. A second conductor came to the first one's aid, and when both of them couldn't come to a clear conclusion, they called for assistance from the next town. The train stopped in mid journey, we waited for several minutes, and two police officers got on board. When I saw the four men bent over the small ticket, softly discussing with each other, "could it really be him? I am never sure with these faces. But an African…" I could not resist and stepped up to them. I was writing my doctoral thesis on precisely that matter of cross-race identification, if I could please have a look at the ominous photograph? When I was shown the ticket, I did not believe my eyes: The photograph depicted the face of a White, reddish-blond man.

Sarah Chiller-Glaus

3. Introduction

3.1. The Situation on the Ground

Verifying someone's identity from a photograph is a very common and not much thought-about task nowadays. On credit cards, passports, train tickets, student cards, driver's licenses, in all these documents we find a photo of the document holder. A large number of occupational areas depend on our abilities to reliably verify the identity of a person by a photograph or video recording. For example, identity parades and identification via video recordings from security cameras are an inherent part of police procedures. Photographs on credit cards and many other identifications (i.e. drivers licence, student ID) should help to verify the bearer of the card. One area of critical importance is the field of border crossings: immigration into or emigration from a state depends on a valid document, and it is vital that the bearer of a document presented at any border crossing be identified correctly. At border control and customs, the security personnel's task is to assess if the photograph in a travel document matches its bearer, and to identify any possible document fraud, e.g., when the bearer travels with the document of a similarly looking double. The state police department of Switzerland requires: "Generally, the owner of a document has to be identified without any doubt"[3] (p. 31).

Figure 3.1: Situation at border control

Our skills in face recognition are indeed truly remarkable, at least as long as we are familiar with a face (e.g. Burton, Wilson, Cowan, & Bruce, 1999, Bruce, Henderson, Newman, & Burton, 2001). Even alterations of hairstyle, external paraphernalia, or the passing of many years do not substantially reduce our ability to recognize familiar faces (Bahrick, Bahrick, & Wittlinger, 1975). The recognition of unfamiliar faces, however, is another issue. It is far from accurate and highly susceptible to image variations, such as viewpoint, lightning or image media (e.g. Hancock, Bruce, & Burton, 2000). Part-based changes such as hairstyle, beards or headdresses have a great influence on the performance in recognizing unfamiliar faces. These findings imply that the job of identity verification where a photograph as old as ten years[4] has to be compared to a person with now different haircut, hair colour, glasses, facial hair, make-up, jewellery, emotional expression and so forth, is in fact not as simple as it seems.

As a matter of fact, the short anecdote at the head of Part II impressively shows the crux of identity verification from photographs: It is highly error-prone. The intention behind the photograph is the prevention of misuse: photographs cannot be forged so easily, and a stolen document cannot be used by an unrightful owner. Yet while theoretically enabling total proof of the rightful ownership, the human ability to verify the identity from the photograph seems at its limit. A striking demonstration of this issue was made by Kemp, Towell, and Pike (1997). In their field experiment, they informed cashiers in a supermarket that a number of shoppers would present them with fraud credit cards depicting another person than the bearer. Despite increased attention to detect the impostors, the cashiers still accepted slightly over 50% of the fraud credit cards as valid. Kemp et al. strongly question the human ability to identify a person by a photograph. Another field of research which addresses the problem of identity verification is the recognition of a person from a survey camera. It is well known that eyewitnesses to crimes often make mistakes at the attempt to identify a face. For example, Bruce, Henderson, Greenwood, Hancock, Burton, and Miller (1999) could show that the matching of unfamiliar faces from CCTV cameras against photographs lead to a high percentage of false responses even when viewpoint and expression where identical between the two images, and worsened even more when they did not match. Equally, when identifying robbers in a fictional bank raid performance was low, even with high quality video footage (Henderson, Bruce, & Burton, 2001). Another study revealed that even skilled personnel such as police officers experienced in forensic identification did not perform better than inexperienced persons when identifying the photograph of a face from the still image of a surveillance camera (Burton et al., 1999). In the light of such evidence, Bruce, Henderson, Newman and Burton (2001) argue that the use of evidence from CCTV cameras in court, where people unfamiliar with the offender are asked to identify the defendant, should be

---

[3] Echt falsch. Leitfaden zur Erkennung von Fälschungen. (2006).

[4] The time span of validity for adults' Swiss passports

avoided, “as resemblances of otherwise unfamiliar faces can be misleading and such judgments are highly prone to error” (p. 217).

Taken together, previous research revealed that we have substantial difficulties when verifying the identity of a person by a photograph. Nevertheless, the task of identity verification by a document is a very common one and conducted at a daily basis at thousands of border crossings, purchases, identity controls etc. world wide. The idea that such a wide spread practice in fact relies on abilities which we do not really possess is intriguing.

3.2. Aim of Research

It is the aim of Part II to shed light on the issue of identity verification by photographs. This shall be done in three chapters, the first one addressing the problems faced, the second possible solutions, and the third methodological considerations arising from the first and second chapter.

In the *first chapter*, the problems of identity verification at border control are assessed, and possible limitations and skills in human performance are systematically analyzed. Experiment 1 addresses the issue of time pressure at border control: Despite high demands on accuracy towards security personnel, reports revealed that in average, a travel document is assessed within approximately ten seconds. This time span includes the verification of the document photograph and a number of other factors of authenticity. Possible influences of time pressure on identity verification are thus discussed. In Experiment 2, the size of the document picture is addressed: A current procedure at border control when doubting the genuineness of a travel document is to enlarge the document photograph and to compare it to a same size photograph of the document bearer. The standard test of identity *before* any doubt has been established, however, is to simply compare the document directly to the bearer without enlargement. It is the aim of Experiment 2 to investigate whether identity verification benefits from enlargement of the photograph. Finally, Experiment 3 addresses the question whether professional expertise of security personnel has an influence on identity verification. The performance of experienced police officers is compared to that of untrained novices. Also, the problem of other-race identification is analyzed, assuming that faces from other races are more difficult to recognize than own-race faces. Generally, the first chapter reveals that performance is poor.

The *second chapter* addresses the question how identity verification performance can be raised. Two experiments are designed to investigate whether training in face recognition leads to better performance. Experiment 4 is conducted with arts students undergoing a course in portrait painting. Their performance is compared to that of novices who have no experience in portraits at all. Experiment 5 contains a fully developed training system on face recognition and shows whether intensive training can help to enhance performance.

The *third chapter,* finally, addresses methodological questions that arise from the research described in the previous two chapters. Experiment 6 addresses the question of matching versus recognition: Many studies on face recognition are based on an old-new recognition paradigm, i.e. the participants first learn a face and later are asked if a certain stimulus belongs to the group of those previously learned. In document verification, however, a different task is faced, namely the matching of two simultaneously presented images. Possible consequences of this setting on identification performance are discussed. Finally, in Experiment 7, the choice of stimuli – i.e. photographs of siblings – is addressed, in particular whether siblings differ from persons not related to each other in terms of their part-based and configural information.

4. Defining the Problem

4.1. Experiment 1: Display Duration

Introduction

At border control and customs, the security personnel's task is to assess if the photograph in a travel document matches its bearer, and to identify any possible document fraud, e.g., when the bearer travels with the document of a similarly looking double. A constantly rising number of passengers at international border crossings require fast and smooth processing at document control desks. This requirement, together with substantial costs for any delay of departure, leads to considerable time pressure at border control. At the same time, the costs for careless security checks remain high. Individual reports from security personnel at border crossings revealed that they have usually only a few seconds of time to assess travel documents. This includes the verification of the document photograph plus a number of other factors of authenticity. It is evident from this setting that there remains no more time for the verification of a photograph than a few short seconds. To examine any possible effect of limited decision time on document verification, this study was conducted. A second aim concerns the verification of anecdotal evidence from different border crossings about an unorthodox practice: As a number of security inspectors revealed, the inversion of the document photograph together with an up-to-date picture of the passenger is supposed to facilitate identification – a procedure which utterly contradicts theoretical findings in face recognition: Yin (1969) was the first to describe that upside-down faces are disproportionately more difficult to recognize than other inverted objects, a finding which is generally referred to as *face inversion effect*. Since then, the inversion effect was replicated in many studies (e.g. Bartlett & Searcy, 1993; Ellis, 1975; Valentine, 1988) and became one of the most important criteria for the distinction between face- and object recognition (e.g. Diamond & Carey, 1986; Farah, Tanaka, & Drain, 1995), as well as a tool to explore what makes faces special (e.g. Farah, Drain, & Tanaka, 1998; Rhodes, Brake, & Atkinson, 1993; Young, Hellawell, & Hay, 1987). Several theories were established to explain the inversion effect and the underlying mechanisms (for a review, see Schwaninger, Carbon, & Leder, 2003). With such strong evidence for the face inversion effect, it seems very implausible why inversion of a travel document together with a photograph taken from the passenger should increase identity verification performance. There remains, however, the possibility that face recognition in an old-new recognition task, as used in most studies on the face inversion effect, and identity verification from a photograph, which includes the matching of two simultaneously presented pictures, involve different processes. It might well be that inversion in the former is detrimental, while for the latter it proves beneficial. Should these presumptions turn out to be true, it would not only be important for application, but also an astonishing theoretical finding.

Methods

Participants

Seventy-two undergraduate students (43 female) aged between 19 and 63 years (M = 24) from the University of Zurich participated in the experiment in exchange for course credit. All had normal or corrected to normal vision and were naïve as to the purpose of this study.

Materials

As stimuli, up-to-date color photographs of 20 pairs of siblings (ten female) aged 16 to 55 were used, depicting the face in frontal view with neutral expression. In addition, their valid document photograph (either passport or identity card, validity up to ten years) were used[5]. Using document photographs provided a challenging test, since a face is likely to change during the validity period of a document. The variance in appearance and age of the stimuli revealed obvious potential for difficulty in document verification. Trials were created as follows: From each pair of siblings, consisting each of person A and person B, four combinations were generated so that both the up-to-date photograph of person A and B were displayed once next to their own document photograph, and once next to their sibling's (see Figure 4.1). The stimuli were presented on a black background, covering approximately 12° of the visual angle in height.

Procedure

The participants' task was to decide whether the two pictures presented depicted the same person or not (simultaneous same-different matching task). Instructions on the procedure were given in written form on the monitor. Response was given by a mouse pressing one of two buttons on the screen labeled "same" or "different". To examine effects of time pressure at border control, the display duration of the stimuli was varied as follows: one second, four seconds, and self-paced. Participants were randomly assigned to one of these three conditions

---

[5] The author would like to thank Timna Tal for the stimulus library which she, equipped with photo camera and scanner, painstakingly collected in numerous visits to the subjects.

(24 participants to each group). All trials were presented twice, once upright and once inverted, resulting in a total of 160 trials (20 pairs * 4 combinations * 2 orientations). All varying factors (orientation, combination of persons A and B, gender of siblings, document type) were counterbalanced across participants using a latin square design. The independent variables were display duration (one second, four seconds, self-paced) and orientation (upright, inverted). The dependent variables were the hit and false alarm rate (a hit was defined as the correct identification of a fraud document, a false alarm as wrong identification of a correct document as fraud). In addition, the participants' rating of confidence about their response on a slider from “unsure” to “sure” on a 90 point scale were analyzed. Participants did not know the ratio of same and different trials, and did not receive any feedback on the correctness of their responses.

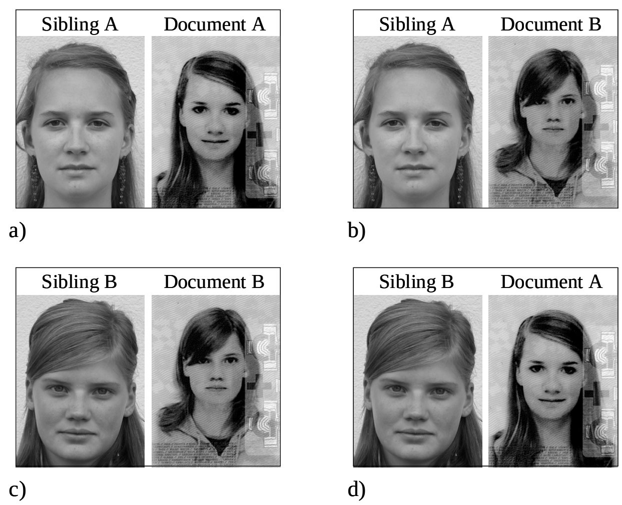


Figure 4.1: Four combinations of up-to-date photograph and document photograph. a) and c) represent “same”-trials, b) and d) “different”-trials.

Results

To calculate identity verification performance, signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 1991) was used. d' was calculated by the formula d' = z(hit rate) – z(false alarm rate). Results of the mean d' values of upright and inverted photographs for the three time conditions are shown in Figure 4.2.

Overall detection performance was relatively low, which confirms the findings of previous studies on face recognition from photographs (see introduction of Part II). A two-way ANOVA with the within-participant factor orientation (upright and inverted) and the between-participant factor display duration (one second, four seconds, and self-paced) revealed a main effect of both factors (orientation: $F(1, 69) = 92.64$, $p < 0.001$, $\eta^2 = .57$; display duration: $F(2, 69) = 25.89$, $p < 0.001$, $\eta^2 = .43$). No interaction was found between the two factors ($F(2, 69) = 0.18$, $p = .18$, $\eta^2 = .05$), which indicates that the relative impact of orientation did not differ between the three display duration conditions.

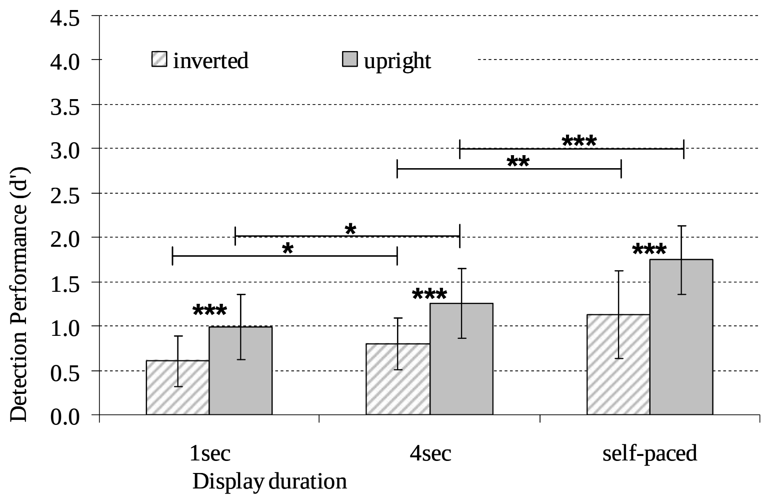


Figure 4.2: Influence of orientation and display duration on mean identity verification performance (d'). Error bars represent standard deviations. * = p < .05, ** = p < .01, *** = p < .001

Paired sample t-tests (two-tailed) pointed out that the higher d' scores for upright faces were mainly the result of a higher hit rate, i.e. more correct identifications of fraud in the upright than in the inverted condition, while the false alarm rate – incorrect "different" responses – did not change considerably between the two orientations (see Table 4.1). In other words: The correct rejections (correct "same" response) were not influenced by orientation, but inverted stimuli evoked more frauds to be missed than upright stimuli.

	1sec	4sec	self-paced
Hit	2.13 *	0.58	1.99
FA	3.59 **	4.55 ***	9.05 ***

Table 4.1: t-values for the difference in detection performance (Hit and False Alarm Rate = FA) between upright and inverted stimuli. * = $p < .05$, ** = $p < .01$, *** = $p < .001$

Identity verification performance significantly rose with increasing display duration, which stresses the importance of a working environment free of time pressure at border control.

Also, in all three display duration conditions, performance was better for upright than inverted stimuli, mirroring a classical face inversion effect. This raises the question why then, in some cases, inversion of the document was reported to facilitate identity verification. To test the possibility that these reports were only the result of increased subjective confidence rather than objectively measurable performance enhancement, a two-factor ANOVA (duration and orientation) of the participants' confidence ratings was conducted. A main effect of orientation ($F(1, 69) = 8.46$, $p < .01$, $\eta^2 = 0.11$) revealed that participants were more confident with their judgment about upright faces, ruling out the possibility of increased confidence in the inverted condition.

Discussion

The results confirm previous evidence that identity verification from photographs is to a high degree error-prone. Even without time pressure, d' for upright faces did not exceed 1.75. Bearing in mind that the recognition of a face is a very common task, the relatively low performance in identity verification of faces is striking.

The results also show that time pressure has a negative effect on identity verification performance: d' drops significantly when collected under a restricted display duration of four seconds, and is reduced even further in the condition of one second display duration. This could be critical since usually only a few seconds are used for the verification of the whole travel document including the picture of the person.

Regarding the inversion of the photographs, the results show that in average, turning the document upside down does not improve, but in fact reduce performance. These results were to be expected when considering the wealth of research on the face inversion effect (e.g. Valentine, 1988; Yin, 1969). They are, however, in contrast to the claims of individual security officers according to which identification is easier from inverted documents. A subjective feeling of increased confidence with inverted faces as a possible explanation for this practice could be ruled out by analyzing the confidence ratings. It therefore remains unclear why the inversion of documents is believed to help identification at all.

Nevertheless, the results do not rule out the possibility than in selected cases inversion of a face provides additional information. On an individual level, roughly 7% of the participants performed contrary to the majority and showed better results for inverted photographs. Further research is needed to establish whether this result is simply a measurement error in this experiment, or whether such behavior represents a stable trait in some people worthwhile to be identified and assisted. Furthermore, performance on the individual level revealed large differences among participants, especially in the conditions where display duration was limited. Such large standard deviations among participants might imply individually developed matching abilities or the use of particular strategies, especially since the internal consistency of the test is rather high (Cronbach's $\alpha = .82$), thus ruling out the possibility that these large standard deviations are due to an unlucky choice of items. If such abilities or strategies can be found consistently, it would be beneficial to develop pre-employment assessment tests to select candidates who are well-suited for the document verification task.

Identity verification is a highly relevant task at thousands of border crossings world-wide. The results found in this study suggest that identity verification using picture ID documents might be more difficult than many people expected. Further obstacles to the task, such as the small size of the photograph in travel documents, or problems arising from cross-race identification, are addressed in the next two experiments.

4.2. Experiment 2: Document Size

Introduction

As revealed by security personnel at border control to the author, when doubting the genuineness of a travel document, the following procedure of identity verification is executed: The passenger is taken a picture of which as accurately as possible matches the details of the photograph in the travel document, such as viewing angle, lightning, facial expression, hair style, visibility of ears, etc. Both photographs – the new one plus the document's – are then displayed next to each other at a screen and compared to each other by a police officer. There is no restriction in time for this comparison. Note that the size of the two photographs is identical and much larger than in the travel document. This practice implies that correct identification of the passenger is facilitated by two factors: first, by enlargement of the photograph to full screen size, rather than directly assessing the small photograph in the document; and second, by identical size of the two images – document photograph and face of the passenger –, a situation which is never given at standard passport control where the viewing angle of the passengers' faces differ constantly from the pictures in the document. The standard test of identity *before* any doubt has been established, however, is to simply compare the document directly to the bearer without enlargement. To test whether the size of the document photograph plays a critical role in identity verification, namely whether enlargement of the photographs would enhance performance, Experiment 2 was conducted. Note that this experiment is of explorative nature, inspired by practical experience at border control, and not based on a theoretical background. The results from this experiment are compared to those of Experiment 1 with identical size of the two photographs. This comparison should allow measuring the influence of image size on identity verification performance.

Methods

Participants

Twenty-two undergraduate students from the University of Zurich participated in the experiment in exchange for course credit. The group consisted of 17 females and five males, aged between 19 and 47 years (M = 22). All had normal or corrected to normal vision and were naïve to the purpose of the study.

As control group, the data of the 24 participant in Experiment 1 (display duration: four seconds) were used.

Materials and Procedure

The same material as in Experiment 1 was used, with the only difference that the document photographs were not of the same size as the up-to-date color photographs, but of a smaller size with a viewing angle as would occur on the retina when looking at a standard Swiss passport photograph from a distance of about 40 cm (2.4° in width and 3.2° in height). The procedure was identical to the one applied in Experiment 1. Display duration was set at four seconds.

Results and Discussion

As in Experiment 1, signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 1991) was used to calculate identity verification performance. A hit was defined as the correct identification of a fraud document, a false alarm as wrong identification of a correct document as fraud. d' was calculated by the formula d' = z(hit rate) – z(false alarm rate). Four participants declared knowing one of the target faces personally; therefore all trials of this pair of siblings were excluded from the calculations. Results of the mean d' values of upright and inverted photographs for both small and large photographs are shown in Figure 4.3.

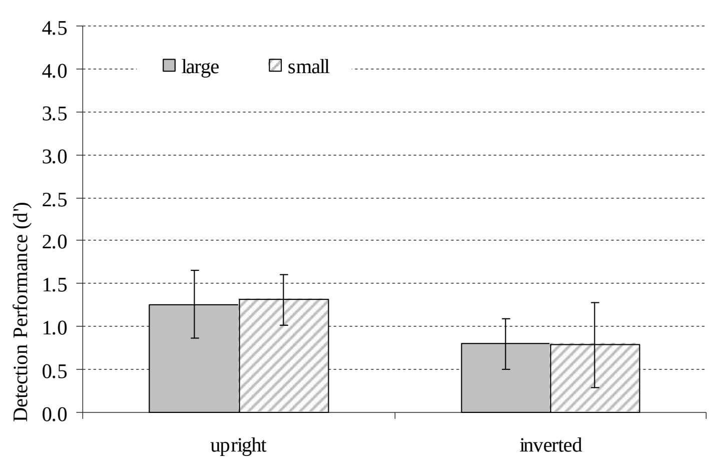


Figure 4.3: Influence of photograph size and orientation on mean identity verification performance (d'). Error bars represent standard deviations. * = p < .05, ** = p < .01, *** = p < .001

A two-way ANOVA with the within-participants factor orientation (upright, inverted) and the between-participants factor size (large, small) revealed that there was no main effect of the photograph's size on identity verification performance ($F(1, 44) = 0.02$, $p = .90$, $\eta^2 = .00$). In fact, there was hardly any difference between large and small photographs, as can be seen from the high p-value of .90. Neither was there a significant interaction between the factors orientation and size (F(1, 44) = 0.12, p = .73, η^2 = .00). As in experiment 1, a highly significant effect of orientation was found (F(1, 44) = 55.10, p < .000, η^2 = .56), which is consistent with previous findings on the face inversion effect (e.g. Yin, 1969).

Regarding document inspection at border control, the obtained results are reassuring: Apparently, the size of the photograph assessed is of no critical importance to identity verification. In a study using CCTV cameras, Liu, Seetzen, Burton, and Chaudhuri (2003) could show that face recognition is robust even with incongruent image resolution. Experiment 2 now showed that this is also the case with incongruent image size. For standard document inspection therefore no implications have to be drawn. The question remains, however, why in case of doubt such fastidious measures are taken to as accurately as possible match the newly taken photograph to the one in the document, including identical size. A possible answer for this circumstance might be the fact that in the experiment, a time limit of four second was administered to the participants for verification, while in reality police officers face no such restriction. Possibly, for fast routine inspection a small photograph in standard travel documents might be sufficient, as is suggested by the results obtained in this experiment. In more difficult cases, however, facilitating measures such as unlimited time and identical photograph size are required for accurate identification. Further research is needed to clarify this point.

4.3. Experiments 3a-c: Expertise and Other-Race Identification

Introduction

Burton et al. (1999) could show that even skilled personnel such as police officers experienced in forensic identification did not perform better than inexperienced persons when identifying the photograph of a face from the still image of a surveillance camera. This finding is particularly intriguing when considering that security personnel usually have several years of experience with the identification of faces and therefore should be expected to possess considerable skills in the task. Indeed, Diamond and Carey (1986) could show that experienced dog breeders performed more accurately in recognizing dog pictures than untrained novices. In the case of security personnel, however, no such effect of experience occurred. One might argue that due to life long exposition to faces from childhood on and the high social relevance of faces, performance already is at ceiling level and cannot be boosted any further by professional experience. This is, however, contrary to the findings described above where identity verification performance is far from ceiling level. The role of expertise, therefore, remains unclear. Although well researched in old-new recognition paradigms (e.g. Diamond & Carey, 1986), it has not yet been investigated in the context of identity verification where the simultaneous matching of two photographs is involved. The nature of expertise in a document control setting was thus addressed by this study.

Another question arises from the fact that document control usually takes place at international sites such as border crossings. In addition to difficulties by image variations as described above, identity verification is further hampered by the fact that faces from other races are more difficult to recognize than own-race faces (Chance, Goldstein, & McBride, 1975; Goldstein & Chance, 1980; Hayward, Rhodes, & Schwaninger, 2008; Malpass & Kravitz, 1969, for a review see Meissner & Brigham, 2001). In other words, faces from other races "all look the same". Of course, this so called *other-race bias* (also other-race effect, cross-race effect) constitutes a considerable obstacle for identity verification in situations where the observer is from another race than the person to be identified. Research on eyewitness testimony, for example, indicates that identification of other-race criminals was highly inaccurate (Doyle, 2001; Wright, Boyd, & Tredoux, 2003); in the case of border control, biased security personnel might be prone to accept fraud documents of other-race passengers as valid.

Although the other-race bias has been found by a large number of studies, it is yet unclear why it occurs. Three possible explanations shall be discussed briefly in the following:

Social Factors: According to the *contact hypothesis* by Chance, Goldstein and McBride (1975) the most important factor for the occurrence of the other-race effect is the degree of quantitative and qualitative contact to other-race members. According to this hypothesis, we become experts at discriminating own-race faces due to continuous contact to members of our own race, while at the same time we do not develop any expertise for other-race faces. The contact hypothesis has been confirmed by some studies (e.g., Rhodes, Tan, Brake, & Taylor, 1989; Wright, Boyd, & Tredoux, 2003), and confuted by others (Chiroro & Valentine, 1995; Ng & Lindsay, 1994). On the grounds of such conflicting empirical evidence, the contact hypothesis has been reinterpreted in the sense that not the actual amount of contact to the other race predicts recognition accuracy, but the social relevance of the contact. For example, White fans of the Black US basket ball team did not exhibit an other-race bias (Li, Dunning, & Malpass, 1988). Prejudices towards the other race, however, do not seem to have any effect on the other-race bias (Ferguson, Rhodes, Lee, & Sriram, 2001). Since the other-race bias is known to play such a hazardous role in eyewitness testimony, several studies have aimed at finding factors which allow an *ex post* diagnose on the accuracy of the testimony, such as confidence or decision time. Again, the results are controversial: Decision time (Smith, Lindsay and Pryke, 2000) was found to be a “postdictor” (see p.542), as was confidence, albeit only for own-race faces (Wright, Boyd, and Tredoux, 2001). Smith, Lindsay, Pryke, and Dysart (2001) however, could not confirm these findings.

Differences in Encoding: Another explanation for the other-race bias might be the differential encoding of configural information in own-race and other-race faces: Rhodes, Tan, Brake, and Taylor (1989) suggested that our face-processing system is more sensitive to configural information in own-race faces than in other-race faces. As was discussed earlier, the recognition of faces is strongly impaired by inversion. Rhodes et al. found that own-race faces evoked a greater inversion effect than other-race faces, suggesting that the recognition of own-race faces relied on configural information more strongly than that of other-race faces. While some authors confirmed these findings (Sangrigoli & de Schonen, 2004; Tanaka, Kiefer, & Bukach, 2004), or even suggested that not only configural information was processed better in own-race faces but also component information (Hayward et al., 2008), others, however, obtained opposite results (Valentine & Bruce, 1986) or equal effects for own- and other-race faces (Buckhout & Regan, 1988). So while most studies suggest that differences in encoding might indeed pose a valuable explanation for the occurrence of the other-race bias, opposing results still remain.

Perceptual Learning: A third hypothesis to explain the processing of faces is the *schema hypothesis* (Vernon, 1955). According to this hypothesis, constant exposure to upright faces over many years leads to the development of a schema for upright faces, implying great expertise in face recognition. This expertise, in turn, is attained at the expense of flexibility. This lack of flexibility would explain reduced recognition performance of faces from unfamiliar races, or of faces in uncommon orientations (i.e. inverted faces). Several development studies have supported the view that a face schema develops over years during childhood (for reviews see Carey, 1992; Ellis, 1992; Johnston & Ellis, 1995). Sangrioli and de Schonen (2004) suggest that this process takes place during the first three years (but see Schwarzer & Leder, 2003 for slightly different results and a more detailed discussion). Furthermore, Diamond and Carey (1986) could show that inversion proved to be particularly detrimental to objects of expertise: Dog experts were significantly better at recognizing upright pictures of dogs than inverted pictures, while no such effect was found for non-experts. Dog experts thus exhibited a larger inversion effect for their objects of expertise when compared to non-experts. Keeping in mind that inversion leads to the disruption of configural information, the schema hypothesis can be interpreted in terms that objects of expertise are processed to a great extent on the basis of their configurations, and that sensitivity to configural information is acquired over the years for objects of expertise.

Taken together, the occurrence of the other-race bias might be the result of many factors: Because of missing social contact, we do not develop a face schema for members of races foreign to us, and therefore process their facial components and configurations differently. At border control, this might imply that other-race passengers are not as easily identified as local passengers, with possibly dire consequences for efficiency and security. On the other hand, the contact hypothesis discussed above implies that frequent contact with the other-race reduces the other-race bias, which might as well apply for security personnel at international border crossings. But since there are confuting results regarding the contact hypothesis (Ng & Lindsay, 1994), it remains unclear whether to expect security personnel to perform well on other-race faces or not. Contrary to the number of studies on the other-race bias in eyewitness testimony, no study has yet been conducted in a document control environment. This question was addressed in Experiment 3.

An additional factor investigated in this experiment was once more inversion: the inversion effect being more pronounced in areas of expertise (Diamond & Carey, 1986), performance on inverted faces can therefore be taken as indicator of expertise. The use of stimuli in both upright and inverted orientation furthermore allows a more detailed analysis of the separate processing of configural and part-based information.

In summary, three factors were tested: First, the question whether expertise in document verification has an influence on identity verification performance. For this, the performance of novices, passport inspectors at border control, and members of a special investigation task force were compared. Second, the influence of the race of a document holder on identity verification performance. And third, the effect of inversion upon this task in order to analyze the influence of expertise upon the processing of configural and part-based information in other-race faces in more detail.

Experiment 3a

This experiment was designed to investigate the effect of expertise upon identity verification of own- and other-race document photographs, for both upright and inverted documents.

Methods

Materials

Thirty-two frontal view color photographs of Asian faces (16 pairs of same-sex siblings, 8 female) plus their document picture, and 32 frontal view color photographs of Caucasian faces (16 pairs of same-sex siblings, 8 female) plus their document pictures were used as stimuli. Photographs were taken in daylight using a digital camera and stored in jpg-format in high quality with a resolution of 300x400 dpi. All Portraits were cut out in Adobe Photoshop 6.0 from the top of the head to the chin. The document pictures were scanned from the valid document belonging to the persons photographed, and stored in jpg-format. Since in the original documents some photographs were in colour and others black and white, all colour information was discarded in the scans. The Stimuli were displayed on a 17’’ TFT screen at a distance of about 55 cm using a custom made software running on DELL Optiplex GX280 computers with Windows XP. Screen resolution was set at 786 x 1024 pixels, with 24-bit colours. All stimuli were 300 pixels in width and 400 pixels in height and subtended a visual angle of about 6.8 degrees in width and 9 degrees in height.

Participants

The group of novices comprised 16 Caucasian undergraduate students (eight female), aged between 22 and 37 years (M = 29), from the University of Zurich who participated in the experiment in exchange for course credit. All reported normal or corrected to normal vision and were naïve to the purpose of the study.

In the group of experts, 24 police officers (two female) aged between 25 and 53 years (M = 34 years) working at document control desks at international border crossings participated in the experiment. Their experience on the job ranged from 2 to 12 years (M = 5 years). All had vision according to the requirements of the State Police Department and were naïve to the purpose of the study.

Procedure

Photographs of same-sex siblings were used as stimuli; each pair of siblings consisted of person A and person B, with two images of each person (photograph and document picture).

From each pair, four combinations were generated so that both person A and B were displayed once next to their own document picture, and once next to their sibling's (for an example of Asian stimuli, see Figure 4.4. An example of Caucasian stimuli is displayed in Figure 4.1, experiment 1). From these four combinations, two versions were created, one in which both photographs were presented upright, and one in which both were inverted. To avoid undue repetition of stimuli, the resulting eight possible combinations per pair (four permutations of photograph and document picture, each in two orientations) were split randomly in two equivalent sets, under the restriction that no combination was present in both upright and inverted orientation within one set. To guarantee highest possible randomization of the stimuli, for each two participants, a new pair of sets was created by splitting the eight possible combinations anew. The resulting sets consisted of a total of 128 trials (16 pairs * 2 races * 4 combinations, two of which upright, and two inverted). Participants were assigned randomly to one of the sets. The participants' task was to decide whether the two pictures depicted the same person or not (simultaneous same-different matching task). Instructions were given in written form on the monitor. The participants gave their response by pressing one of two buttons labeled "same" or "different" on the screen using the mouse, and rated the confidence of their choice with a slider (from "unsure" to "sure") on a 90 point scale. Since individual reports from security personnel at border crossings revealed that they usually have only a few seconds of time to assess travel documents, a time limit of four seconds for identity verification was administered. Participants did not know the ratio of same and different trials, and did not receive feedback on the correctness of their responses.

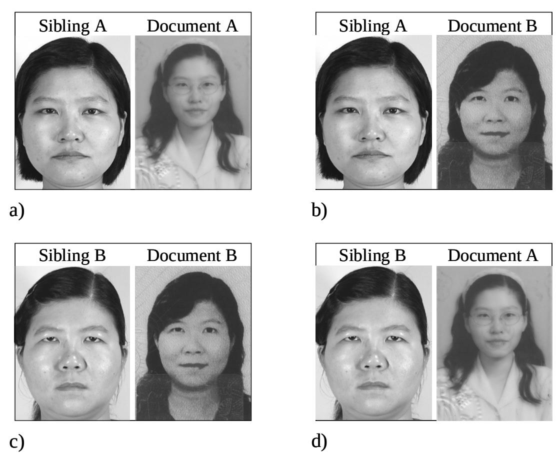
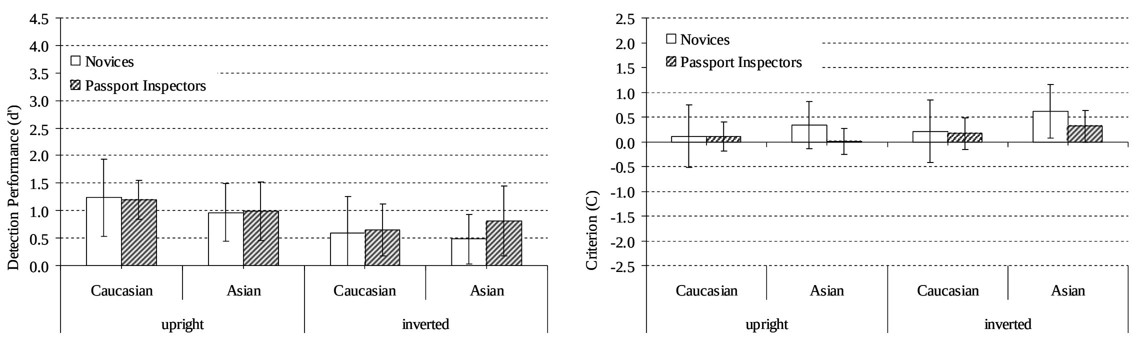


Figure 4.4: Four combinations of up-to-date photograph and document photograph. a) and c) represent "same"-trials, b) and d) "different"-trials.

Results and Discussion

Signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 1991) was used to calculate identity verification performance. A hit was defined as the correct identification of a fraud document, a false alarm as wrong identification of a correct document as fraud. d’ was calculated by the formula d’ = z(hit rate) – z(false alarm rate). Results of the mean d’ values of upright and inverted photographs for both novices and border controllers are shown in Figure 4.5a. In addition to d’, the values of the criterion, respectively the response bias (C) was calculated by the formula C = 0.5 * (– z(false alarm rate) + z(hit rate)). Positive values are associated with a tendency towards lenient responses (e.g. pressing the *same*-button more often); values close to zero indicate unbiased behaviour. Results for the mean C values are displayed in Figure 4.5b.

Two participants declared knowing one of the targets personally, therefore all trials of this pair of siblings were excluded from the calculations of these two participants’ data set.



a) b)

Figure 4.5: a) Average identity verification performance d’, and b) criterion C of novices and passport inspectors for both Asian and Caucasian faces in upright and inverted orientation. Display duration of stimuli: four seconds. Error bars indicate standard deviations.

As can be seen in Figure 4.5a, overall detection performance was relatively low, which confirms the findings of previous studies that identity verification from photographs is not accurate (Kemp et al, 1997). For both the d’- and C-values, a separate three-way ANOVA with the within-participant factor orientation (upright, inverted) and race (Asian, Caucasian), and the between-participant factor expertise (novices, passport inspectors) was conducted to analyze the influence of the race of the passport holder on one hand, and inversion of the document on the other hand, on identity verification performance (d’) and response bias (C) of both novices and experienced passport inspectors. The three factors shall be discussed separately in the following.

Expertise: The analysis of both d' and C revealed no significant main effect of expertise (d': $F(1, 38) = 1.07$, $p = .31$, $\eta^2 = .03$; C: $F(1, 38) = 2.00$, $p = .17$, $\eta^2 = .05$). These results indicate that passport inspectors did not perform substantially better than novices in the task of identity verification, nor did their response tendency (bias) differ significantly. These results on the performance of security personnel in face recognition are in line with Burton et al. (1999) who found that even experienced police officers could not identify the photographs of a face from the still image of a surveillance camera. But when keeping in mind that the passport inspectors who took part in this experiment had several years of experience on the job, such lack of skill is surprising nevertheless.

Race: No significant main effect of race was found for d' ($F(1, 38) = 1.35$, $p = .25$, $\eta^2 = .03$), nor was there a significant interaction between race and any other factor. This finding seems contrary to the other-race effect, according to which Asian faces should be more difficult to identify than Caucasian faces. It might be that the general performance for both Caucasian and Asian faces was simply too low for any effect of race to occur. The analysis of C, however, revealed a significant main effect of race ($F(1, 38) = 15.91$, $p < .001$, $\eta^2 = .30$), along with a significant interaction between race and expertise ($F(1, 38) = 11.41$, $p < .01$, $\eta^2 = .23$). Despite the fact that the actual identity verification performance did not differ between Asian and Caucasian faces, these results indicate that Asian faces systematically lead to a response bias: Apparently, a relatively low false alarm rate was achieved at the expense of a high miss rate, meaning that participants perceived Asian faces to look more similar than Caucasian faces and therefore often missed to correctly identify Asian passports as fraud. Moreover, a three-way ANOVA (expertise, race, and orientation as factors) of the participants' confidence ratings revealed that they were more confident about Asian faces than Caucasian faces (main effect of race, $F(1, 38) = 8.38$, $p < .01$, $\eta^2 = .18$), which again supports the idea that Asian faces are perceived more homogenous and therefore subjectively easier to identify as identical. These findings are in line with other studies according to which other-race faces seem to look more similar to each other than own-race faces (see introduction of Part II). Concerning the significant interaction between expertise and race for the criterion C, the data indicate that experienced passport inspectors are significantly less biased about Asian faces than novices. This might be explained by the contact hypothesis (Chance et al., 1975) according to which frequent contact with other races helps to effectively reduce the other-race bias. Passport controllers at international border crossings are likely to get sufficient contact with Asian passengers for their responses to be significantly less biased.

Orientation: For both d' and C a significant main effect of orientation was found (d': $F(1, 38) = 31.87$, $p < .001$, $\eta^2 = .46$; C: $F(1, 38) = 16.41$, $p < .001$, $\eta^2 = .30$). For C, the interaction between orientation and race was also significant ($F(1, 38) = 9.89$, $p < .01$, $\eta^2 = .21$), all other interactions for C and d' were not. The significant main effect of both measures support previous findings on the face inversion effect (see introduction of Part II). Inverted faces appear to be more difficult to identify than upright faces, which can be seen from the higher d'-values for upright faces. Additionally, they evoke a bias towards more lenient response criteria, explaining their higher C-values. Inverted faces, thus, seem to appear more similar than upright faces. As in Experiment 1, these results raise the question why, in some cases, inversion of the document was reported to facilitate identity verification. Again, the three-way ANOVA (expertise, race and orientation) of the participants' confidence ratings revealed that participants were more confident with their judgment about upright faces (main effect of orientation: $F(1, 38) = 37.15$, $p < .001$, $\eta^2 = .49$), ruling out the possibility of increased confidence in the inverted condition. Taken together, no evidence was obtained in favor of the unusual practice to turn around the document for easier identification. Regarding the other-race bias, the results imply the following: As mentioned in the introduction of Part II, other-race faces seem to be encoded differently from own-race faces, most likely in the sense that their configural information cannot be processed as readily (e.g. Rhodes et al., 1989). When considering that inversion of a face mainly renders its configural information inaccessible, the face inversion effect then should be less pronounced for other-race faces. This is, however, not the case in Experiment 3, since the interaction between race and orientation was not significant. It therefore seems likely that both races are processed similarly in terms of their configurations and separate parts. Whether this is a general finding – which would then of course be in opposition to the aforementioned evidence on differences in encoding – or whether this effect is a result of the matching paradigm in this experiment, remains subject to further research.

In summary, the results gathered in Experiment 3a revealed that Asian faces were perceived more homogenous than Caucasian faces, that they were processed similarly in terms of their configural and part-based information, and that inversion of the document did not help identity verification. The most striking finding, however, was the fact that despite long years of practice at border control, passport inspectors did not perform better than novices when identifying the passport of a passenger as genuine or as double. When asked about their performance after the experiment, most participants reported being short of time when assessing the photographs. The time limit of four seconds was originally administered to mirror time pressure at border control, where travel documents were assessed within only a few seconds of time. In order to avoid any unrealistic data due to overly harsh time restrictions, however, the experiment was replicated with display duration of ten seconds. Results are described in Experiment 3b.

Experiment 3b

Since it was derived from Experiment 3a that the display duration of four seconds for each trial was not sufficient for reliable identity verification, display duration was raised to ten seconds in Experiment 3b. Apart from this change, Experiments 3a and 3b were identical.

Methods

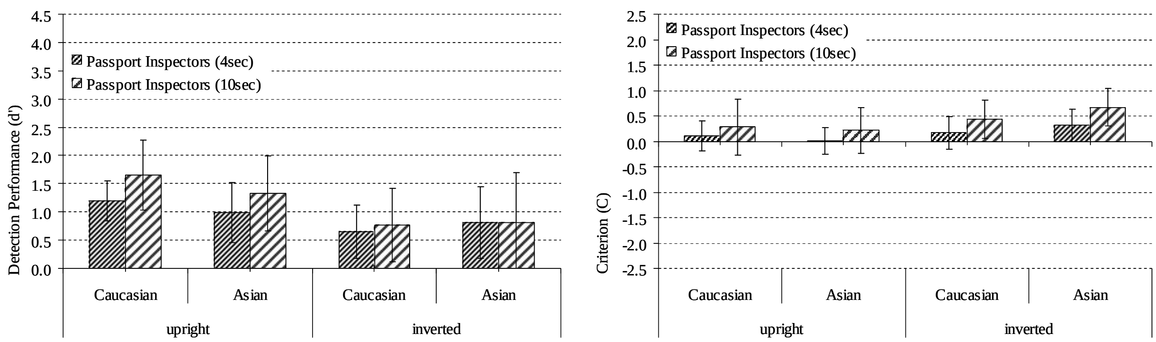
Materials and Procedure were identical to Experiment 3a, with the only difference that the display duration of the stimuli was set at ten seconds rather than four seconds. If the participants wanted to reply faster, they could click away the stimulus at any time.

Sixteen police officers (one female) aged between 29 and 61 years (*M* = 37.5 years) of the same division as in Experiment 3a (not the same participants) took part in the experiment. Their working experience on the job ranged from one to 14 years (*M* = 8 years). All had vision according to the requirement of the State Police Department and were naïve to the purpose of the study.

Results and Discussion

The results of Experiment 3b (display duration: ten seconds) are compared to the identity verification performance of the police officers in Experiment 3a (display duration: four seconds). Since the participants were free to click away the stimuli before the full ten seconds elapsed, the reaction times were analyzed to make sure that the two experiments really differed in terms of time used for decision. The comparison of average reaction time from onset of stimulus display until clicking of answer button (5.69 seconds for four seconds display duration, 8.59 seconds for ten seconds display duration) revealed that the participants took almost three seconds longer when allowed ten seconds than when restricted to four. The average reaction times also show that the participants did not use the full ten seconds, but on average made their decision before the full time elapsed. A three-way ANOVA (display duration, race, and orientation as factors) for reaction time revealed a significant main effect of display duration ($F(1, 38) = 63.71$, $p < .001$, $\eta^2 = .63$), indicting that the two experiments differ significantly from each other in terms of display duration and therefore can be subjected to further analysis. The ANOVA for reaction times furthermore revealed a significant main effect of race ($F(1,38) = 11.83$, $p < .01$, $\eta^2 = .24$) in the direction of Caucasian faces generally requiring longer reaction times than Asian faces. No other main effect and no interactions were significant.

As described in Experiment 3a, signal detection theory was used to analyze the data. The measures for detection performance (d') and criterion (C) were subjected to a three-way ANOVA with the factors display duration (four seconds, ten seconds), race (Asian, Caucasian) and orientation (upright, inverted). The results of d' and C are displayed in Figure 4.6.



a) b)

Figure 4.6: a) Average identity verification performance d', and b) average criterion C of passport inspectors at four and ten seconds display duration for both Asian and Caucasian faces in upright and inverted orientation. Error bars indicate standard deviations.

The benefit of a longer display duration narrowly missed significance, as was revealed by the analysis of d' (no main effect of display duration, $F(1, 38) = 3.70$, $p = .06$, $\eta^2 = .09$). For the criterion (C), a main effect of display duration was found ($F(1, 38) = 8.93$, $p < .01$, $\eta^2 = .19$) indicating that a longer display duration significantly increased the response bias. For both measures, no other-race bias occurred (no main effect of race, d': $F(1, 38) = 0.69$, $p = .41$, $\eta^2 = .02$, C: $F(1, 38) = 1.62$, $p = .21$, $\eta^2 = .04$), but a significant interaction of race and orientation (d': $F(1, 38) = 4.60$, $p < .05$, $\eta^2 = .11$, C: $F(1, 38) = 9.03$, $p < .01$, $\eta^2 = .19$). For d', the difference between Asian and Caucasian faces is diminished by inversion, meaning that when inverted, faces cannot be distinguished reliably anymore, no matter what race. As for C, inversion seems to enhance the difference between Asian and Caucasian faces, meaning that the response bias is equal for both races when upright, but larger for Asian faces when inverted. Both results point to the direction that other-race faces are more difficult to process, be it in actual identity verification performance, or in response bias. The analysis of both d' and C revealed a significant main effect of orientation (d': $F(1, 38) = 43.94$, $p < .001$, $\eta^2 = .54$, C: $F(1, 38) = 22.49$, $p < .001$, $\eta^2 = .37$), confirming the face inversion effect. No other interaction was significant.

The main aim of Experiment 3b was to find out whether identity verification performance increased when given more time, as was suggested by the participants of Experiment 3a. No such effect was found, although significance was missed only narrowly. It might well be that with a larger sample a significant effect of display duration could have been obtained. The three-way ANOVA of the participants' confidence ratings revealed that they were more confident with longer display duration ($F(1, 38) = 8.28$, $p < .01$, $\eta^2 = .18$), which is consistent with the comments of the participants in Experiment 3a requiring more time for reliable identification. But regardless of this tendency towards better identification, what seemed most striking was that overall identity verification still remained poor. Experiment 3b shows yet another sample of experienced passport inspectors, and the display duration was increased to a level higher than actually required, but still d' reached only a maximum of 1.65 for upright Caucasian faces. Such limited performance raised the question whether the task of identity verification from photographs simply exceeds human face recognition abilities and thus cannot be performed reliably, independently of any experience on the job. This interpretation would be in line with the studies mentioned in the introduction of Part II about very limited skills in identity verification, even of experienced police officers. Another interpretation, however, concerns the nature of expertise: The sample of passport inspectors was chosen for this study in virtue of their long experience with document verification; but they had not undergone any form of skills assessment or special training for face recognition during their employment as security personnel. As was shown in Experiment 3a, they did not perform significantly better than novices. The question thus remains whether expertise in identity verification can be obtained by practice alone. In order to rule out the possibility that the wrong sample of experts was chosen for this study, Experiment 3c was conducted with a different sample of highly qualified security personnel: investigators.

Experiment 3c

Since it was unclear whether the experience of the passport inspectors in Experiment 3a and 3b was enough to qualify them as experts in identity verification, members of a highly skilled investigators task force within the police department were chosen for experiment 3c. These police officers' daily job was to locate criminals with the help of wanted photos. All reported a natural disposition for face recognition, had undergone a detailed admission procedure including a practical training before full acceptance into the task force, and had afterwards received specific training in investigation. As in Experiment 3b, display duration of the stimuli was set at ten seconds to account for the feedback given by passport inspectors in Experiment 3a about four seconds not being sufficient for reliable identification. This setting, however, does not allow direct comparison between members of the special investigating task force and novices, since the latter were tested under a display duration of four seconds (Experiment 3a). The same experiment was therefore conducted with another sample of novices under a display duration of ten seconds. By comparing the results of all three samples of this study, namely novices, passport inspectors and investigators, the results of this experiment should contribute to a more detailed understanding of the role of expertise.

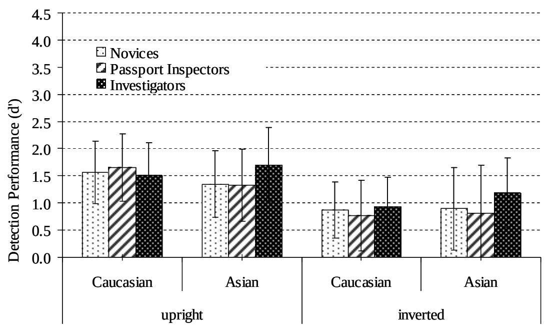
Methods

Materials and Procedure were identical to Experiment 3b (display duration of ten seconds). Sixteen investigators of a special task force (all male), aged between 33 and 55 years (M = 42) took part in the experiment. Their experience ranged from 1 to 15 years (M = 6). All had vision according to the requirement of the State Police Department and were naïve to the purpose of the study.

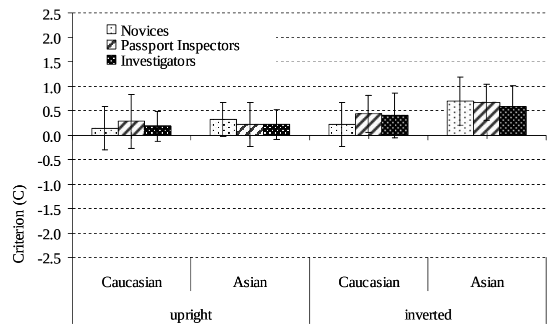
In the control group of novices, 20 Caucasian undergraduate students (15 female), aged between 18 and 37 years (M = 19 years), from the University of Zurich participated in the experiment in exchange for course credit. All reported normal or corrected to normal vision and were naïve to the purpose of the study.

Results and Discussion

The results of the investigators and novices of Experiment 3c are compared to the passport inspectors of experiment 3b. As in all experiments so far, signal detection theory was used to analyze our data. The measures for detection performance (d') and criterion (C) were subjected to a three-way ANOVA with the factors occupation (novices, passport inspectors, special investigating task force), race (Asian, Caucasian) and orientation (upright, inverted). The results of d' and C are displayed in Figure 4.7.



a)



b)

Figure 4.7: a) Average identity verification performance d', and b) average criterion C of novices, passport inspectors, and investigators at ten seconds display duration for both Asian and Caucasian faces in upright and inverted orientation. Error bars indicate standard deviations.

The most striking result is that there was no significant effect of expertise, neither for d' (F(1, 53) = 0.99, p = .38, η^2 = .04) nor for C (F(1, 53) = 0.20, p = .82, η^2 = .01). Nor were there any significant interactions with the factor expertise for d'. No single comparison between the three samples, in all four conditions (Caucasian-upright, Asian-upright, Caucasian-inverted, Asian-inverted), was significant. This result indicates that neither investigators nor passport inspectors perform better in identity verification than untrained novices. Regarding the security personnel's high aptitude and professional experience, such findings are intriguing.

The interpretation lies near that the verification of a person's identity from a photograph is simply too difficult a task for the human eye. There is, however, an alternative explanation to this phenomenon: As Riegelnig & Schwaninger (2006) could show in a study with X-Ray screeners, the ability to detect prohibited items in a passenger's bag declined with increasing age of the screener. For experience, the contrary is true, i.e. any special skills are expected to increase with the passing of time. Since these two factors are unavoidably confounded, a possible effect of expertise in Experiment 3a-c might not be detected. To test this possibility, a partial correlation between the work experience of the security personnel (both passport controllers and investigators, display duration four and ten seconds) and their detection performance d' was conducted and corrected for age. A scatter plot with the security personnel's d' values and experience of years is displayed in Figure 4.8.

The analysis, however, revealed no significant effect of experience (r(57) = .12, p (two-tailed) = .18). Increasing age and with it a reduction of cognitive abilities can therefore not be responsible for the fact that even with long experience identification performance does not increase. It seems that greater skills in identity verification are simply not given, and cannot be obtained through practice.

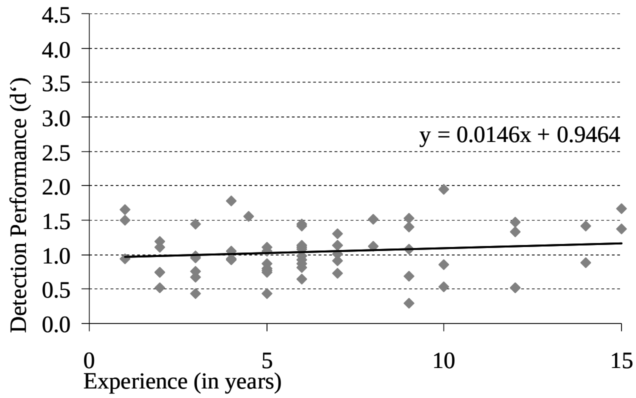


Figure 4.8: Correlation between work experience and detection performance (d') of security personnel (investigators and passport inspectors, display duration 4 and 10 seconds).

Regarding the factors race and orientation, the ANOVA of d' and C (Figure 4.7a and b) revealed the following: For d', there was no significant effect of race ($F(1, 53) = 0.00$, $p = .98$, $\eta^2 = .00$), but C significantly differed for Asian and Caucasian stimuli ($F(1, 53) = 17.32$, $p < .001$, $\eta^2 = .25$), Asian faces generally leading to a larger response bias than Caucasian faces. For C, also the interaction between race and orientation was significant ($F(1, 53) = 8.02$, $p < .01$, $\eta^2 = .13$), but not for d'. So, albeit not evident in actual performance, there was again a "hidden" other-race effect in the sense that Asian faces lead to a larger response bias than Caucasian faces. Also, novices seemed to be more biased about other-race faces than security personnel, as can be derived from the significant interaction between race and expertise for C ($F(1, 53) = 3.58$, $p < .05$, $\eta^2 = .12$). As in Experiments 3a and 3b there was a significant effect of orientation for both measures (d': $F(1, 53) = 68.57$, $p < .001$, $\eta^2 = .57$, C: $F(1, 53) = 34.42$, $p < .001$, $\eta^2 = .39$), which again is in line with the face inversion effect. No other interactions were significant.

General Discussion

In three experiments, the identity verification performance of two different groups of experts, namely passport inspectors working at border control and investigators of a special task force, was compared to that of novices. As stimuli, document photographs of Caucasians and Asians were used, both in upright and inverted orientation. The results confirm previous evidence that identity verification from photographs is to a high degree error-prone. Even without time pressure, d’ for upright Caucasian faces did in none of the samples exceed 1.65, performance for inverted or Asian stimuli generally being even lower. Bearing in mind that the recognition of a face is a very common task, such limited performance in identity verification of faces is striking.

The results also show that even long years of experience in document control do not substantially raise identity verification performance: neither experienced passport inspectors, nor highly qualified members of an investigating task force performed better than untrained students. The results confirm findings from previous research on identity verification (Bruce et al., 1999; Hancock et al., 2000; Kemp et al., 1997). These studies covered a wide range of occupational areas in which the identification of a person’ identity was required, e.g. eyewitness testimony, verification of credit cards, or identification from video recordings. In all of them, performance proved to be far from accurate, be it for untrained laymen or for experienced police officers. A plausible explanation for this is that on their daily jobs security personnel never get a feedback on their identification skills: It seems unlikely that an impostor reveals his true identity after successfully hoodwinking the passport inspector. Also, the assumption lies near that the task of identity verification from photographs simply overtaxes human abilities. The prevalent position on face recognition holds that humans are experts in face recognition (e.g. Diamond & Carey, 1986; Schwaninger et al., 2003 for a review). Yet while we indeed might possess exceptional abilities in the memory of faces, the verification of identity in two simultaneously presented images – with sometimes substantial difference between the images (different external features, lightning, age, etc.) – seems to be more difficult. This study, together with the afore mentioned research on the topic, indicate that face recognition in an old-new recognition paradigm where the differences between the stimuli mostly lie in subtle changes of small details in the otherwise unchanged image, and identity verification from two photographs taken on separate days, sometimes even years, are two separate mechanisms which cannot be compared to one another, and that great skills in the former do not necessarily lead to high performance in the latter.

A second aim of this study was to investigate the other-race effect. Contrary to previous research (e.g. Malpass & Kravitz, 1969; Meissner & Brigham, 2001), no such effect was found, i.e. the identity of Asian faces was no more difficult to verify than of Caucasian faces – or, Caucasian faces already were so difficult to identify that the factor race was of no consequence anymore. It is only in the criterion that a difference between the two races was found, Asian faces evoking more lenient response behavior than Caucasian faces. For the experts in this study, such a finding might be explained by the contact hypothesis as proposed by Chance et al. (1975): Frequent contact with members of the other race is supposed to reduce the other-race bias, and it is plausible that at international border crossings the passport controllers and investigators had enough contact with Asian passengers for the other-race bias in the criterion to disappear. This line of argument, however, does not account for the fact that for novices too, no significant other-race effect was found. Nor does it explain why there was no significant interaction between race and expertise. A possible explanation for the lack of such an effect might be that the design of this study differed from other studies in the sense that most previous research used an old-new recognition paradigm (Meissner & Brigham, 2001), while here a simultaneous matching task was used. The findings of this study are very robust throughout all our experiments, regardless of the profession of the participants. The author therefore comes to the same conclusion as drawn already from the low baseline performance, namely that face recognition as researched in previous studies, and the simultaneous matching of two images taken at two separate dates as used in the experiments here, require different mechanisms that cannot be compared to one another. Apparently, while memory of faces is largely affected by race, the direct comparison between two simultaneously presented images is not. Whether this is the case because of other-race faces being processed equally well as own-race faces when matched simultaneously, or because baseline performance is already so low that the race of a face does not make much of a difference anymore, remains to be analyzed.

Regarding the inversion of the photographs, the results show that in average, turning the documents upside down does not improve, but in fact reduce performance. It remains unclear why the inversion of documents is believed to help identification at all. However, as was already discussed in Experiment 1, the results do not rule out the possibility that in selected cases inversion of a face provides additional information. As in Experiment 1, there was a number of participants also in Experiment 3 (13%) who performed contrary to the majority and showed better results for inverted photographs. If further research can show that there are indeed persons with a stable trait or special technique which allows them to perform better than the average when given access to inverted photographs, it might be worthwhile to develop pre-employment assessment tests to select candidates who are well-suited for the document verification task.

Taken together, this chapter showed that identity verification at border control is at its limits. Not only was the overall performance proven to be very low, but also experience in the task was not able to raise performance to a reliable level. It is the aim of the next chapter to identify ways by which to improve performance.

5. Towards a Solution

As was shown in Experiments 1-3, identity verification from photographs is error-prone. Baseline performance was low in the first place, but even experience in the task did not seem to raise performance significantly: Experienced passport inspectors, and even investigators of a specially trained task force did not perform better than novices.

The question remains whether good performance in identity verification is based on experience or natural disposition alone, as was the case in Experiment 3a and b (passport inspectors) and 3c (investigators), or whether it can be learned and trained. For this purpose, two experiments were conducted. Experiment 4 was carried out with students of the Zurich University of Arts who had enrolled in a portrait painting course. The aim of this experiment was to investigate whether their performance after the course was better than before, which could be taken as a sign that identity verification was indeed trainable. Experiment 5 then involved the development of a real training system for face recognition, with stimuli of increasing difficulty according to the participants' performance, over the time span of one week. Taken together, the findings of both experiments should contribute to the understanding of how – if at all – identification from photographs can be learned and what measured have to be taken for maximal increase in performance.

5.1. Experiment 4: Arts Students

Introduction

As mentioned in Part I of this work, the information contained in faces is commonly classified in component and configural information. One characteristic attribute of face recognition is the face inversion effect, meaning that the processing of faces is highly orientation-dependent (Yin, 1969). The face inversion effect is mainly due to the disruption of configural information when faces are inverted, while the processing of component information remains relatively unimpaired. Diamond and Carey (1986) could show that for objects of expertise, such as dog pictures for experienced dog breeders, inversion had an even more detrimental effect than for common objects. As mentioned above, the participants of Experiment 4 were students of the Zurich University of Arts before and after a portrait painting course. Regarding their gain in face perception skills, a possible conclusion from this theoretical background could be that increased performance in identity verification would go in hand with a larger inversion effect. To test this possibility, the faces used to test identity verification ability were shown both in upright and inverted orientation. This design allowed a detailed analysis of the process of learning, particularly on the involvement of configural and component information in identity verification.

As an additional test to the effect of training upon performance, again both Caucasian and Asian faces were used in the experiment: The class of students comprised only Caucasian members, and the curriculum of the course did not indicate that other-race faces were subject to painting. It could therefore be assumed that a possible training effect would only occur with own-race faces. If, however, training would extend also to Asian faces, this could be taken as a sign that the training of one race could be transferred to another. This could be of particular interest to security personnel at border control, namely that the other-race bias could be reduced by training with one race alone.

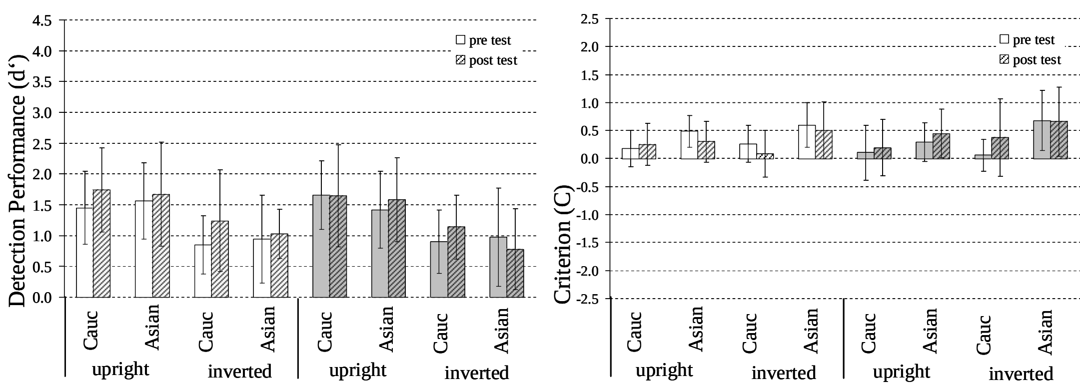
Taken together, three factors were tested in Experiment 4: First and foremost the effect of training upon performance. To do so, performance of art students was measured pre- and post-training. Their performance was compared to a control group which was tested twice, in the same interval as the arts students. The second factor investigated was orientation with the aim to gain a deeper understanding of the involvement of configural and component information in training. And third, the factor race was investigated to test whether skills in the identification of one race could be transferred to another race.

Methods

Materials and Procedure were identical to Experiment 3b (display duration of ten seconds). In the experimental group, 12 Caucasian students of the Zurich University of the Arts (9 female), aged between 21 and 40 years (*M* = 24) took part in the experiment. All were enrolled in a portrait painting course, taking seven weeks to complete. The first testing session took part before the course, the second testing session seven weeks later after completion of the course. The control group comprised 16 undergraduate students of the University of Zurich (13 female), aged between 18 and 37 years (*M* = 21.5), who took part in the experiment for course credit. None of them reported having any particular experience in face recognition and reported normal or corrected to normal vision.

Results and Discussion

Signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 1991) was used to calculate identity verification performance. A hit was defined as the correct identification of a fraud document, a false alarm as wrong identification of a correct document as fraud. d' was calculated by the formula d' = z(hit rate) – z(false alarm rate). Results of the mean pre-training and post-training d' values for art students and control group are shown in Figure 5.1a. In addition to d', we calculated the values of the criterion, respectively the response bias (C) by the formula C = 0.5 * (– z(false alarm rate) + z(hit rate)). Positive values are associated with a tendency towards lenient responses (e.g. pressing the *same*-button more often); values close to zero indicate unbiased behaviour. Results for the mean C values are displayed in Figure 5.1b.



a) b)

Figure 5.1: a) Average identity verification performance d', and b) average criterion C of arts students (white) and control group (grey) for both pre test and post test. Error bars represent standard deviations. Cauc = Caucasian stimuli, Asian = Asian stimuli.

Two participants declared knowing one of the targets personally, therefore all trials of this pair of siblings were excluded from the calculations of these two participants' data sets.

Before subjecting the data to any deeper analysis, the pre-test performance of art students was compared to that of the control group to ensure the validity of the control group. T-Tests (two-tailed) for all the four conditions (Caucasian-upright, Caucasian-inverted, Asian-upright, and Asian-inverted) revealed no significant difference between the two samples, indicating that the arts students did not have any superior abilities in face processing arising from their field of study, and therefore the group of non-arts students could function as control group to the art students.

A three-way ANOVA was conducted with the factors training (pre-training, post-training), orientation (upright, inverted), and race (Asian, Caucasian) for the art students. The analysis of d' revealed a significant main effect of training ($F(1, 11) = 7.31$, $p < .05$, $\eta^2 = .40$), with higher scores for post-training than pre-training. The same ANOVA for the control group did not reveal such an effect ($F(1, 15) = 0.19$, $p = .67$, $\eta^2 = .01$). These results indicate that performance in identity verification can significantly be increased by training. When analyzing the effect of training in a four-way ANOVA for the combined data sets of the arts students and the control group (fourth factor: sample, i.e. arts students versus control group), however, the effect of training narrowly missed significance ($F(1, 26) = 3.36$, $p = .08$, $\eta^2 = .11$), and there was no significant interaction between the factors training and sample ($F(1, 26) = 1.36$, $p = .26$, $\eta^2 = .05$). Also for the criterion, no significant main effect and no interaction were found (main effect of training: $F(1, 26) = 0.10$, $p = .76$, $\eta^2 = .00$, interaction: $F(1, 26) = 3.86$, $p = .06$, $\eta^2 = .13$). These data indicate that the effect of training – although found when analyzing the two samples separately – is rather small. A possible reason for that might be that the course in portrait painting was unspecific regarding the students' artistic approach towards faces; also, only seven weeks of course work might be rather little time to make the students true experts. Nevertheless, as was shown from the difference between the arts students and the control group, there is at least some evidence that training can be suitable for raising performance in identity verification.

The four-way ANOVA furthermore revealed a significant main effect of orientation for both d' and C (d': $F(1, 26) = 40.01$, $p < .001$, $\eta^2 = .61$, C: $F(1, 26) = 6.20$, $p < .05$, $\eta^2 = .19$), but no significant interaction between orientation and sample for either measure (d': $F(1, 26) = 0.04$, $p = .85$, $\eta^2 = .00$, C: $F(1, 26) = 1.65$, $p = .21$, $\eta^2 = .06$). Regarding the other-race bias, the difference between Asian and Caucasian faces was not significant for d' ($F(1, 26) = 1.46$, $p = .24$, $\eta^2 = .06$), but highly significant for C ($F(1, 26) = 42.69$, $p < .001$, $\eta^2 = .62$), with responses for Asian faces being more biased. The interaction between race and sample was not significant for either measure (d’: $F(1, 26) = 0.86$, $p = .36$, $\eta^2 = .03$, C: $F(1, 26) = 0.41$, $p = .53$, $\eta^2 = .02$). For both d’ and C, no other interaction was significant. These data reveal similar findings as obtained in Experiment 3: Inversion of the photographs does not help, but rather hinder identity verification, which is consistent with the large amount of research on the face inversion effect. Regarding the missing interaction between the samples and the factor orientation, the following can be assumed: Diamond and Carey (1986) showed that in fields of expertise, inversion reduces recognition performance to a stronger degree than in other fields. The fact that the arts students did not differ from the control group regarding their behaviour with inverted faces indicates that – while nevertheless increasing identity verification to a certain degree – the short training in portrait painting was not sufficient to raise their skills to a level where inversion would be critical. Therefore, no additional conclusions can be drawn about the processing of configural and part-based information. As for the other-race effect, the results are in line with the findings from Experiments 3a-c: In performance, no other-race effect was found. It seems that either the task of simultaneous matching evokes similar answers for both Asian and Caucasian faces and is not comparable to other tasks (e.g. old-new recognition), or identity verification performance is not high enough to discriminate between own- and other-race faces in the first place. As in Experiments 3a and c, however, there was a “hidden” other-race effect of the criterion in the sense that the participants were more biased with other-race faces than with faces from their own-race, which is at least partially consistent with previous literature on the other-race effect. The missing interaction between race and the two samples can be interpreted along the same line as was the case with orientation, namely that the training of the arts students was not enough to significantly raise identity verification performance of own-race faces above those of other races. The alternative interpretation of a transfer of knowledge about own-race faces gained during the course to other-race faces (which were not subject to training) – although very interesting from a theoretical point of view – seems unlikely regarding the fact that also for the control group, there was no interaction between training and race ($F(1,15) = 0.47$, $p = .50$, $\eta^2 = .03$).

Taken together, it seems that training has no specific effect on the orientation of the faces, nor on their race. But nevertheless, there is evidence that training might indeed be an appropriate measure to raise identity verification performance. The effect in Experiment 4 was rather small, which is not surprising when considering the fact that the training was not specifically designed for the task, but merely the participation of a portrait painting class. In order to find out if identity verification could be helped by a sophisticated, standardized training program, Experiment 5 was conducted.

5.2. Experiment 5: Training

Introduction

The aim of Experiment 5 was to investigate whether the ability to identify faces from photographs can be enhanced by training. For this endeavor a special training system was developed by which the participants could train their abilities with levels of increasing difficulty. A test before and after training allowed to measure the direct influence of training on performance.

The challenges faced at test construction were manifold. First, there was the problem of stimuli: The task set to the participants – as in previous experiments – was to be to decide whether two simultaneously presented photographs depicted the same person or not. In the same-condition, however, the two images could not be identical, lest the participants relied on simple picture matching which would have been too easy a task. Therefore, of each person two images had to be taken on separate days, so as to allow natural changes in expression, hairstyle, and external paraphernalia to occur between the two shooting session. The lookout for such a face library of sufficient scale proved to be difficult enough.

A further challenge was to create a set of stimuli of increasing difficulty, confronting the participants with ever more demanding images, and at the same time to control all the factors manipulated in the faces in a standardized way. The solution to this challenge involved the morphing of faces, which indeed bears a number of great advantages: First and foremost, the stimuli thus obtained are highly realistic looking images. Other than in experiments where only single features are replaced, or the spatial distances altered for creation of a different identity, morphing gradually changes the whole face in a way which at all points remains natural to look at. Second, morphing provides an elegant solution to the problem of how to create stimuli of increasing difficulty: The ratio of each of the two faces morphed together can be controlled easily and thus provides a gliding scale of images of increasing distance from the original. The higher the percentage of another identity morphed into the original photograph, the easier it should be to conclude that it is not the original identity anymore.

Last but not least, another challenge involved motivational aspects: As training was to be maintained over a time span of one week, the task had to be interesting so as not to unduly strain the participants' patience. For this, the training system was set up as a computer game.

Methods

Materials

128 frontal view color photographs with neutral expression and even illumination of 64 Caucasian faces (two photographs from each person taken on separate days), half female, from the *AR face database* (http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html) were used as stimuli. The use of two photographs from separate shooting sessions allowed realistic testing of face recognition abilities without relying on simple picture matching. From each of these 64 individuals, two new images were created with morphing technique using Abrosoft FantaMorph software (www.fantamorph.com), resulting in a total of four images per individual: a) the photograph of the first shooting session; b) the photograph of the second shooting session; c) a morph of the first and second shooting session, still depicting the same individual, although morphed to a fictional image ("identical morph", see Figure 5.2a); and d) a morph of the second session with another individual of identical sex and similar hairstyle, resulting in a new fictional identity ("non-identical morph", see Figure 5.2b). From these four images, two combinations were created: shooting session 1 with the identical morph ("same-trial", see Figure 5.2c), or shooting session 1 with the non-identical morph ("different-trial", see Figure 5.2d).

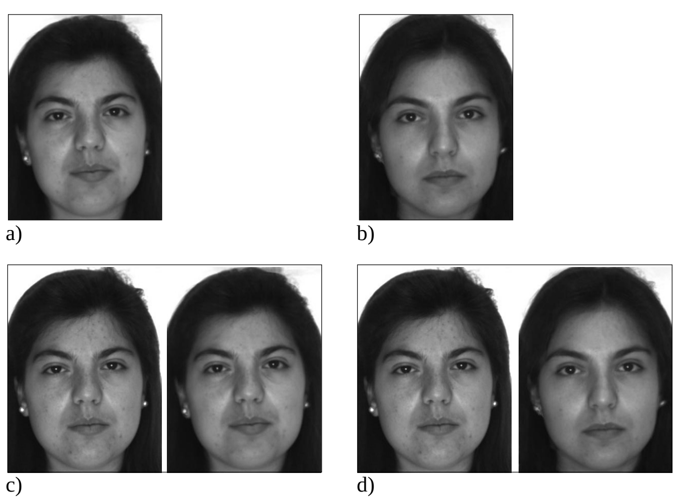


Figure 5.2: Example of Stimuli used in the training. a) morph of two photographs of one individual ("identical morph"), b) morph of two photographs of two individuals ("non-identical morph"), c) same-trial (= shooting session 1 with identical morph), d) different-trial (shooting session 1 with non-identical morph). Note that all the morphs in this example contain 50% of each photograph, which corresponds to level 1 in the training.

Of each of the 64 same-trials and 64 different-trials, five versions were created with different morphing ratios of 50, 45, 40, 35, and 30%, the percentages always indicating the ratio of the alternative photograph morphed into shooting session 2. In total, 640 trials were created[6] (64 individuals * 2 combinations (same/different) * 5 morphing ratios).

Participants

16 participants (8 female), all with a high-school or university degree, aged between 17 and 32 (M = 26) participated in the experiment. All were naïve to the purpose of the study and had normal or corrected to normal vision.

Procedure

Test: To assess the participants' ability to identify faces from photographs, the same test as described in Experiments 3b and c was used, once before and once after training. Display Duration was set at ten seconds. Comparison between the two sessions allowed measuring the influence of training upon identity verification performance.

Training: The training for identity verification was set up as a game with five levels of difficulty[7]. It was the participants' task to assess whether two simultaneously presented photographs depicted the same person or not (simultaneous same-different-matching task, see Figure 5.2c and Figure 5.2d for an example) by clicking on a green button labeled "same" or on a red button labeled "different" with the mouse. For every correct answer they received a yellow smiley, a wrong answer resulted in loss of all the smileys obtained so far. At the collection of eight smileys (the number of trials that cannot be obtained by guessing within a confidence level of 5%), the participants rose to the next higher level. They also rose one level if they had seen all 128 trials within one level without getting eight answers right in a row. The levels of difficulty were defined by the morphing ratio of the images (50% = easiest, 30% = most difficult). Instructions were given in written form on the monitor. The participants underwent training for one week with one training session per day, with a break on two separate days of their choice (five training sessions in total).

---

[6] Here the author wishes to express her explicit gratitude to Corinne Frey for the tremendous effort of morphing and editing all the 640 stimuli.

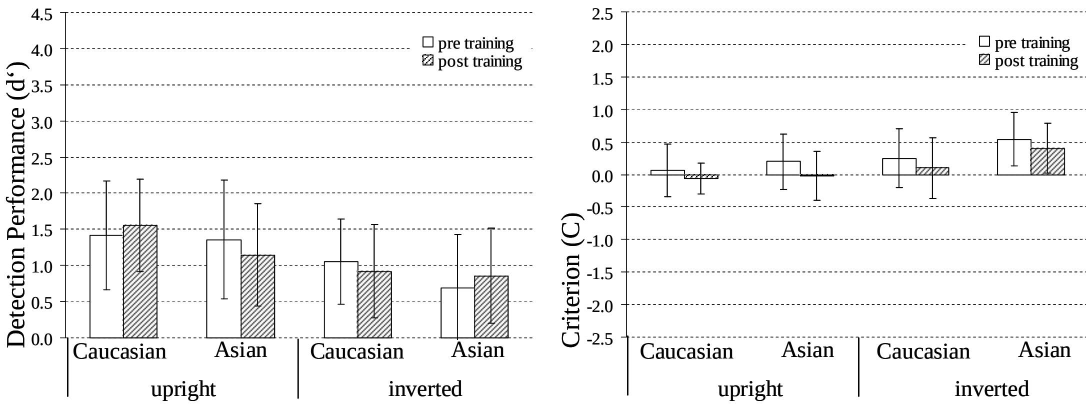
[7] Special thanks go to lic. phil. Markus Ruh for programming the necessary software.

Results

As in the experiments described before, signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 1991) was used to calculate identity verification performance of both test and training. In the test, a hit was defined as the correct identification of a fraud document, a false alarm as wrong identification of a correct document as fraud. Along the same lines, in the training a hit was defined as correct identification of the two photographs as different persons, while a false alarm was defined as wrong assumption that two photographs which in fact depicted the same person were of two different persons. d' was calculated by the formula d' = z(hit rate) – z(false alarm rate), the response bias (C) by the formula C = 0.5 * (–z(false alarm rate) + z(hit rate)). Three participants declared knowing one respectively two of the targets in the test personally, therefore all trials of these pairs of siblings were excluded from the calculations of these three participants' data sets. The results on test and training are discussed separately in the following.

Test

Results of the mean pre-training and post-training d' values are shown in Figure 5.3a, results for the mean C values in Figure 5.3b. A three-way ANOVA with the factors training (pre-test, post-test), orientation (upright, inverted), and race (Asian, Caucasian) was conducted. The analysis revealed no main effect of training for d' (F (1, 15) = 0.02, p = .89, η^2 = .00), but a significant effect for C (F (1, 15) = 6.50, p < .05, η^2 = .30). These results indicate that the training during one week did not raise the participants' performance. The training did, however, influence their response behaviour in the sense that they were less biased towards perceiving the faces to look the same. The analysis furthermore revealed no main effect of race for both d' (F (1, 15) = 2.95, p = .12, η^2 = .16) and C (F (1, 15) = 3.91, p = .07, η^2 = .21). As described in previous experiments, these data indicate that there is no other-race bias, i.e. identity verification performance itself does not differ between Asian and Caucasian faces. The "hidden" other-race effect – a significant difference between Asian and Caucasian faces in the criterion, indicating that responses for Asian faces were more biased – did not occur here, although significance was missed only narrowly. Also as described in previous experiments, there was a main effect of orientation for both d' (F (1, 15) = 11.82, p < .01, η^2 = .44) and C (F (1, 15) = 26.56, p < .001, η^2 = .64), confirming once more the face inversion effect. For either measure, no interactions were significant.

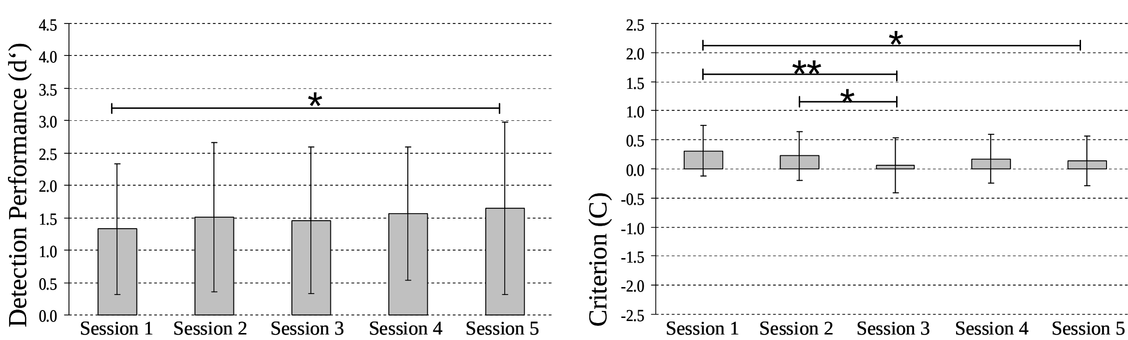


a) b)

Figure 5.3: a) Average identity verification performance d', and b) average criterion C of the test before and after training. Error bars represent standard deviations.

Training

Results of the d' values of the five sessions are shown in Figure 5.4a, those of the C values in Figure 5.4b. Although there is a slight tendency for better performance towards the end of the week, the one-way ANOVA with the factor "training" (i.e. sessions 1-5) revealed no significant effect for d' ($F(4, 60) = 0.65$, $p = .63$, $\eta^2 = .04$). Also for the criterion, there was no significant effect of training ($F(4, 60) = 1.02$, $p = .41$, $\eta^2 = .06$) However, single comparisons between the five sessions revealed a significant difference in the d' values between session 1 and 5 ($t(78) = -1.96$, $p < .05$). All other single comparisons were not significant. These results indicate that the one-week training might have an effect on identity verification performance, although this effect is rather small.



a) b)

Figure 5.4: a) Average identity verification performance d', and b) average criterion C for the five training sessions. Error bars represent standard deviations. * = $p < .05$, ** = $p < .01$, *** = $p < .001$

Discussion

It was the aim of this experiment to raise identity verification performance by training. To do so, the participants underwent training in face identification for the duration of one week, and performed a test to measure their overall identification ability before and after the training. The results of the test revealed no significant effect: Performance after the training was equal to that before. But in the training sessions performance slightly rose towards the end of the week, although this effect was rather small (ANOVA not significant, only single comparisons).

The results obtained in this experiment do not give a clear answer to the question whether skills in identification can be obtained by training. The results of the pre- and post-test indicate that they cannot be obtained, since d’ did not significantly differ between the two tests. However, there might be a methodological explanation to this: In the test, the display duration of the stimuli was set at ten seconds, as was explained in Experiments 3b and c. The fact that the photographs in the test were visible only for a restricted amount of time, while in the training they stayed as long as required, was irritating for several participants. They felt that the skills they had obtained in the training were not those required later in the test. Further research is needed to investigate whether a test with no time limit might produce other results. The analysis of C at least revealed that the training had rendered the participants more critical towards the photographs in the sense that more differences were perceived in the faces after the training.

As for the training, the experiment revealed a tendency towards better abilities at the end of the week, indicating that learning is indeed possible. This effect was only small, but on the other hand, training was only very short. Regarding the amount of training required in other occupational fields, five sessions of 30 minutes each are indeed a very short time to obtain reliable skills. In the field of X-ray screening, for example, the acquisition of true expertise requires much longer: As is stated by Koller, Schwaninger, Michel, & Hardmeier (2008), the security personnel working at baggage screening undergo recurrent training for six months such as to maintain their high standard of screening abilities. When considering these long times of learning, it seems likely that in this experiment, training was simply too short for any large effect to be established. Regarding the positive tendency in this experiment, an appropriate horizon of time might very well allow significant abilities in the identification of faces to develop. Again, further research is needed to answer this question.

6. Methodological Considerations

So far, the problem of identity verification from photographs was addressed in Part II. The stimuli used in the five experiments were faces of siblings, such as to allow a realistic setting of similarly looking doubles. Also, a matching paradigm was used to measure identity verification performance, i.e. the simultaneous presentation of two photographs, which the participants had to assess as either depicting the same person or two different persons.

These methods were chosen to simulate a plausible setting at border control as realistically as possible. However, there are some considerations to be taken into account regarding the methodological correctness of this procedure: Data obtained in a matching paradigm might not be compared to data obtained with another method, and the use of siblings as stimuli might lead to artifacts in the data which are not welcome. In order to rule out the possibility that the findings of Part II are in some way confounded by the methods applied, two control experiments were conducted to ensure the quality of the data. Experiment 6 addresses the research paradigm used for data collection, and Experiment 7 investigates the role of siblings as stimulus material.

6.1. Experiment 6: Matching versus Recognition

Introduction

A classical paradigm to measure face recognition performance is the old-new recognition paradigm, in which participants are presented a set of different faces and in a later stage are asked to indicate for a larger set of faces whether a specific face has been part of the earlier set ("old") or not ("new") (see for example Tanaka & Farah, 1993; Yin, 1969). At border control, however, we encounter a different situation, namely the matching of two simultaneously present images, i.e. the face of the passport holder and the passport photograph. Compared to an old-new recognition task, such a situation might not require the same memory resources, but rely on simple matching of features.

One of the aims of the experiments described so far was to analyze the face inversion effect in a border control setting: As anecdotal evidence suggests, the inversion of the passport photograph together with a picture of the bearer might facilitate identification. So far, no evidence for this could be obtained: In all of Experiments 1 to 5, inversion of the photographs reduced identification performance rather than enhancing it. However, the question remains whether inverted stimuli are processed equally in a matching task and an old-new recognition task: Matching of two simultaneously presented photographs might not require the processing of configural information, but simply rely on part-based information. Reports from participants of previous experiments about comparing single features being the easiest way to verify the identity of a face, together with the police department's instruction manual that security personnel should assess "shape of face, ears, chin and nose; eyebrows, size, moles etc."[8] (p. 31), seem to confirm this hypothesis. But when assuming that in face matching the focus lies on part-based information, and that the main cause for the face inversion effect is the disruption of configural information, a matching paradigm might evoke different results for inverted stimuli than the old-new recognition paradigm, at least concerning the inversion effect's magnitude. In the setting of identity verification chosen in this work, this implies that although a face inversion effect did occur in all the experiments, it might have been even stronger had the data been obtained with an old-new recognition paradigm. So in order to analyze the influence of the task administered in the test on the magnitude of the inversion effect, Experiment 6 was conducted. The stimuli used in this experiment were identical to those used in the matching task in Experiment 1, but here set up in an old-new recognition-

---

[8] Echt falsch. Leitfaden zur Erkennung von Fälschungen. (2006).

paradigm. The comparison of Experiment 6 to Experiment 1 should allow measuring the direct influence of the task on identity verification performance.

Methods

Materials

The stimuli used in this experiment were identical to those of Experiment 1 (see detailed description of the stimuli there): color photographs of 20 pairs of siblings (ten female), plus their valid document photograph. Each pair of siblings consisted of person A and person B, with two pictures of each person (photograph and document picture). For maximal randomization, the pairs were split up anew for each participant into group A and group B, resulting in 24 different sets of person A and B.

Participants

Twenty-four undergraduate students of the University of Zurich (18 female), aged between 18 and 42 years (M = 25.5), took part in the experiment for course credit. They were all naïve to the purpose of the study and reported normal or corrected to normal vision.

Procedure

The experiment consisted of a learning phase and an experimental phase. During the learning phase, the up-to-date photographs of person A of each pair of siblings were sequentially presented in random order in the center of a black screen for ten seconds each. Participants were required to memorize the faces. After one cycle, the photographs were then presented again in the same order.[9] In the experimental phase, the document pictures of all persons A and B were presented. Presentation occurred sequentially, in the center of a black screen. Each document picture was used twice upright and twice inverted[10], resulting in a total of 160 trials (40 document pictures*2 orientations*2 presentations). The order of presentation was randomized with the constraint that identical pictures did not immediately follow one another. Document pictures of person A were target faces, document pictures of person B distractors. The participants' task was to decide whether the document pictures depicted a person previously learned in the learning phase or not. The display duration of each trial was self-paced.

---

[9] In Experiment 1, each photograph was presented twice, once in combination with the document picture of person A, and once with person B. For the sake of comparison between Experiment 1 and 6, pictures were also presented twice in Experiment 6.

Results and Discussion

As in the experiments described before, signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 1991) was used to calculate identity verification performance. In line with Experiment 1, a hit was defined as the correct classification of a face as new, a false alarm as wrong classification as an old face as new. d' was calculated by the formula d' = z(hit rate) – z(false alarm rate). Also the response bias (C) was calculated, by the formula C = 0.5 * (– z(false alarm rate) + z(hit rate)). The results of the d' values are shown in Figure 6.1a, those of C in Figure 6.1b.

To analyze the influence of the task upon identity verification performance in general, and in particular on the magnitude of the face inversion effect, a two-way ANOVA was calculated for both d' and C with the factors task (matching, recognition; between-participants, i.e. Experiment 1 vs. 6) and orientation (upright, inverted, within-participants). The analysis of d' revealed a main effect of orientation ($F(1, 46) = 30.06$, $p < .001$, $\eta^2 = .40$), confirming the face inversion effect. There was, however, no main effect of task ($F(1, 46) = 3.40$, $p = .07$, $\eta^2 = .07$), and no interaction between task and orientation ($F(1, 46) = 0.72$, $p = .40$, $\eta^2 = .02$). For the criterion C, the opposite results occurred: There was a significant main effect of task ($F(1, 46) = 16.10$, $p < .001$, $\eta^2 = .26$), but no effect of orientation ($F(1, 46) = 1.22$, $p = .28$, $\eta^2 = .03$), and no significant interaction ($F(1, 46) = 2.53$, $p = .12$, $\eta^2 = .05$). As can be seen in Figure 6.1a, there is a tendency for the matching task to produce higher values than old-new recognition. This is consistent with reports from participants who declared that the recognition task was exceptionally difficult. This effect, however, is not significant, as the ANOVA revealed. Given the fact that the old-new recognition task indeed was very challenging, involving identity verification by a photograph different from the photograph previously learned, it is surprising that performance was actually well above chance level. On the other hand, it seems extraordinary that the matching task, perceived comparably easy by the participants, did not produce significantly higher performance than the recognition task. On the contrary: When looking at the criterion, the response behaviour in the matching task was significantly more biased compared to old-new recognition in the sense that the participants perceived faces to look more similar, which – regarding the difficulty of the task – would rather be expected for recognition. These results seem to stress our exceptional abilities in face recognition, even under such difficult circumstances as were administered in Experiment 6.

---

[10] Again, for the sake of comparison with Experiment 1, the document pictures were presented twice.

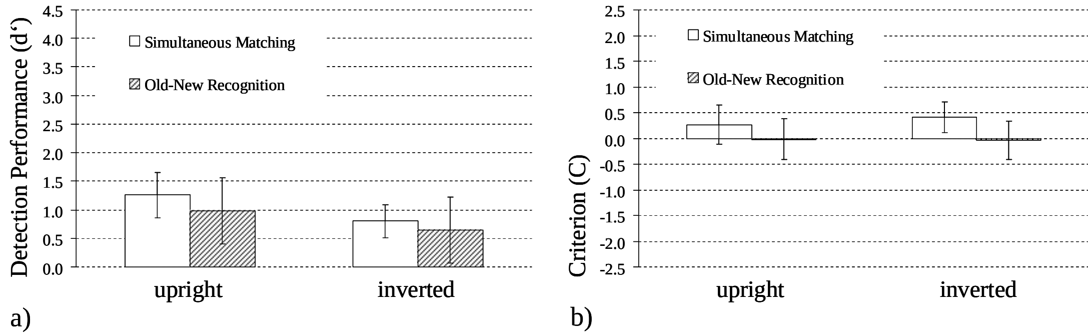


Figure 6.1: a) Average identity verification performance d', and b) average criterion C for simultaneous matching and old-new recognition. Error bars indicate standard deviations.

Experiment 6 was designed to investigate the magnitude of the inversion effect in both an old-new recognition and a simultaneous matching task. As a side effect, it revealed that the difference in overall performance between the two tasks was not substantial. It was argued in Experiments 3a-c that some of the findings on identity verification – e.g. lack of high expertise – resulted from the choice of paradigm. Experiment 7 now indicates that it is not so much the paradigm, but the choice of stimuli which makes identity verification from photographs so difficult: In other studies stimuli are often identical apart form relatively small configural or featural changes (e.g. Barton, Deepak, & Malik, 2003; Diamond & Carey, 1986; Ellison & Massaro, 1997, Ingvalson & Wenger, 2005), while in this study the participants were required to recognize the identity of a person from two different photographs taken years apart. It is evident that the latter setting is much more demanding on our cognitive abilities. Experiment 7 does thus not confute the arguments brought forward in Experiment 3, but adds a new aspect.

Regarding the main research question of Experiment 7, i.e. the magnitude of the inversion effect in the two paradigms, the following can be concluded: Neither for d' nor for C was there a significant interaction between the two factors, indicating that the face inversion effect did occur rather independently of the task, both in relative and in absolute terms. It seems therefore unlikely that in a matching task configural information is of less importance, since inversion of the stimuli reduces performance to the same extent as in the old-new recognition task. Regarding the fact that most participants of the matching experiments in this study reported comparing single features rather than looking at the global face, the findings of this experiment are surprising.

6.2. Experiment 7: Kinship

Introduction

In all the experiments on document verification conducted in this part (Experiments 1-5), the photographs of same-sex siblings were used as stimuli. Siblings provided an easy way to collect a large enough set of similarly looking doubles. From the results obtained in the experiments, conclusions were drawn about the inversion effect and the processing of the faces' separate features and configurations. There is, however, a possible caveat in this design: since the siblings were chosen randomly and not in a standardized procedure, it is unclear if they really look more alike than non-kin faces. Should they not look more alike, no consequences on the interpretation of the data would arise, only that in reality the task of identity verification might be even more difficult with impostors choosing doubles of high resemblance – and, of course, that the highly time-consuming collection of same-sex siblings' photographs would be proven futile. On the other hand, if siblings indeed do look more alike than non-kin pairs of faces, then the question arises what such high likeness is based upon: Should the siblings chosen for these experiments share the same configurations, the inversion effect would be larger than with non-kin doubles (see introduction of Part II on the connection between inversion and configural information). On the other hand, should part-based information be the main source of likeness among the siblings, the inversion effect would be smaller. Either way, the nature of the siblings' resemblance might have an influence on the interpretation of Experiments 1-5. So in order to rule out that the results were partially influenced by the stimulus material, Experiment 7 was conducted.

In this experiment, the configural and part-based information contained in the faces were separated in order to analyze their influence upon recognition independently. For this, a method by Schwaninger, Lobmaier, and Collishaw (2002) was used: Blurring of the stimuli by a Gaussian filter is supposed to reduce part-based information. On the other hand, cutting out single features and then rearranging them anew (scrambling) should destroy configural information. Using this technique, the faces of both siblings and non-kin were used as stimuli in this experiment to analyze the influence of kinship upon the processing of configural and part-based information.

Method

Participants

Eighteen participants (8 female), aged between 23 and 57 years (M = 28) voluntarily took part in the experiment. All were holders of a university degree and had normal or corrected to normal vision.

Materials

Sixteen frontal view color photographs of Caucasian faces (8 pairs of same-sex siblings, 4 female) were chosen from the set of stimuli used in Experiments 1-5. All photographs were processed with Adobe Photoshop, proportionally scaled to the same face width of 300 pixels, cut out with the elliptical marquee tool with a two pixels feather and a fixed size of 280 x 410 pixels and placed on a black background (see Figure 6.2a). Copies of these faces were then used to create blurred, respectively scrambled, versions as follows: For the blurred stimuli, all color information was discarded, and the images were then blurred using a Gaussian filter with a radius of 11 pixels (Figure 6.2b). For the scrambled stimuli, the images were cut into 10 parts (2 eyes, 2 eyebrows, 2 cheeks, nose, chin, mouth and forehead), using the polygonal lasso tool with a 2 pixel feather and then rearranged in four different ways as described by Schwaninger et al. (2002) (see Figure 6.2c). The intact faces represented test faces, the scrambled and blurred faces targets[11].

The sets of trials – one with scrambled, one with blurred stimuli – were composed of all possible combinations for all same-sex stimuli, i.e. every intact face was combined with every same-sex scrambled (respectively blurred) stimulus, resulting in eight possible combinations per stimulus. These combinations were grouped in three separate conditions: The first condition comprised the combination of intact face with its own scrambled (or blurred) version (“identical”). In the second condition, every intact face was combined with its sibling’s scrambled (or blurred) version (“sibling”). Finally, in the third condition, every intact face was combined with the scrambled (or blurred) version of an unrelated face (“non-kin”). Note that the conditions “identical” and “sibling” each contained 16 trials, the condition “non-kin” 96 trials. Examples of the three conditions, once scrambled and once blurred, are shown in Figure 6.2d. All combinations were shown once, resulting in a total of 128 trials (16 stimuli x 8 combinations) for each the scrambled and the blurred set. The Stimuli were displayed on a 17’’ TFT screen using a custom made software running on DELL Optiplex

[11] Special thanks go to Anne Bode for the time-consuming task of creating the stimuli for this experiment.

GX280 computers with Windows XP. Screen resolution was set at 786 x 1024 pixels, with 24-bit colors.

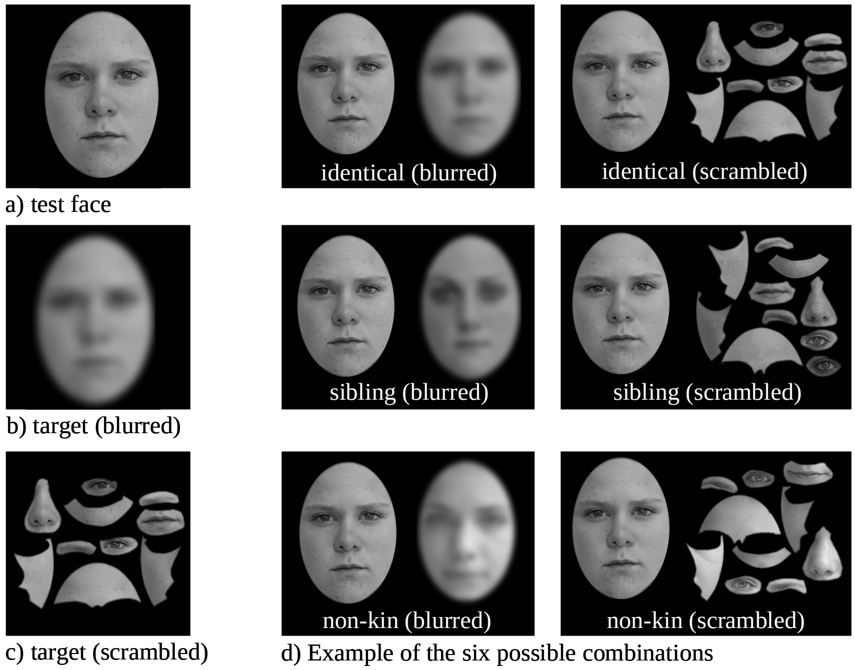


Figure 6.2: Example of Stimuli. Description see in the text.

Procedure

The Participants were randomly assigned to either the scrambled or the blurred condition (8 in the former, 10 in the latter). They were tested one by one; instructions were given in written form on the monitor. The stimuli were presented in a sequential matching paradigm, i.e. the test face was presented for 2000ms, followed by a mask (1000ms), and then by the target face, for as much time as required by the participant. The task was to decide whether the two images depicted the same person or not. The participants answered by pressing buttons labeled “same” or “different” on the screen. The following trial could be started by pressing the space bar. No feedback was provided. After four practice trails, the main experiment started.

Results and Discussion

In this experiment, the percentage of correct responses was taken as measure for identification. The value of 0.5 represents chance performance. As can be seen in Figure 6.3, identical trials were easiest to recognize, followed by non-kin. The identification of siblings proved to be most difficult and close to chance level. In all three conditions, the scrambled stimuli were identified better than the blurred stimuli. A two-way ANOVA with the within-participants factor kinship (identical, sibling, non-kin) and the between-participants factor information type (scramble, blur) revealed a highly significant main effect of kinship ($F(1.51, 24.18) = 63.91$, $p < .001$, $\eta^2 = .80$). The main effect of information type missed significance by a fraction ($F(1, 16) = 4.46$, $p = .05$, $\eta^2 = .22$), most likely it would be significant if more participants were tested. There was no significant interaction between the two factors ($F(1.51, 24.18) = 0.38$, $p = .63$, $\eta^2 = .02$).

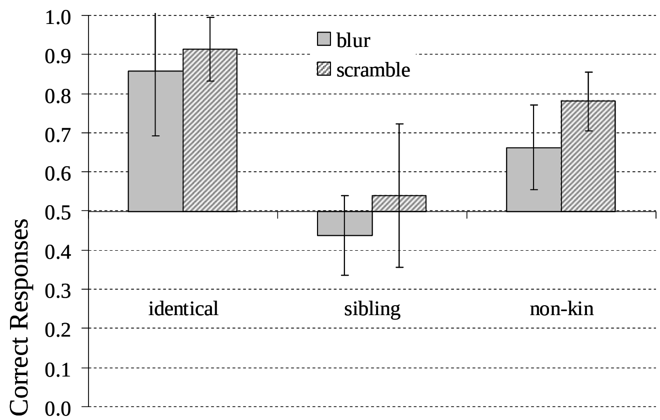


Figure 6.3: Correct responses for identical photographs, siblings, and unrelated persons. Error bars represent standard deviations.

The results indicate that it was very difficult for the participants to distinguish siblings from identical photographs. While identical trials were mostly perceived as identical, and non-kin fairly often as non-kin, siblings were in about half the cases mistakenly identified as identical. The difference between the non-kin and the sibling condition thus shows that siblings were perceived to look more alike to each other than unrelated persons. As for the question how this similarity was achieved, the missing interaction between kinship and type of information revealed that it is neither configural nor part-based information in particular, but rather a combination of both which create the effect. Generally, the loss of part-based information turned out to be more detrimental to identification than of configural information, which can be derived from the higher performance for scrambled stimuli. This is contrary to other findings stressing the importance of configural information (for a review, see Schwaninger et al., 2003; or introduction of Part II). However, it needs to be kept in mind that the number of participants was very small (only 8 participants in the scrambled condition). While regarded sufficient for a control study such as this experiment, the results should be treated with caution and not applied to a general trend too liberally.

It was the aim of this experiment to investigate whether the use of siblings as stimuli might have an unwanted side-effect on the inversion effect described in previous experiment. Since siblings do not appear to be alike in only their parts or their configurations, such objections can be ruled out.

7. Closing Words and Outlook

In this part, a serious problem in border security was addressed: The fact that identity verification form photographs is highly prone to errors. This question was tackled with a number of different stimuli – European and Asian faces, upright and inverted – with a variety of different samples, such as police officers, professional investigators, or arts students.

Generally speaking, the experiments in this part revealed the following patterns: The race of the passport holder did have no influence upon identification. This is contrary to previous findings on the other-race effect. The most likely explanation to such conflicting results might be the fact that the overall performance was very low already, to an extent that the race might have been of no influence anymore. Also, the inversion of the photographs reduced performance drastically, which is consistent with previous research on the face inversion effect. Since in most experiments inversion did not interact with other factors (e.g. race, expertise), the specific processing of configural and part-based information seems independent from those.

The most striking finding, however, was that throughout all the experiments the average identity verification was low, and that experience was not able to raise it. Training tended to be helpful for identification, although the effects were small. It needs to be born in mind that the test administered in the experiments was not thoroughly validated according to test-psychological criteria and might not be of sufficient quality to live up to the high standards set for pre-employment assessments or certification purposes. It gives, however, a good enough impression of the situation to show that the problem is severe. Identity verification from photographs or video recordings is a highly relevant task not only at thousands of border crossings world-wide, but in a wide range of other occupational fields. The fact that security personnel are entrusted with a job which they are in fact not able to reliably perform reveals a serious gap in the maintenance of security.

The results of Part II clearly put into question the use of photographs in documents, at least where human identification is required. It remains to be seen whether further research for the development of a suitable training system could ameliorate the situation. Otherwise, one possibility would be the use of automated identity verification by computers, as is already done at Faro Airport in Portugal, for example[12]. For pictures of lesser quality, however – such

[12] Neue Zürcher Zeitung, Mittwoch, 9. Juli 2008, p. B1

as older photos, out of focus photos or images with high superposition – machines might still not be able to perform accurately, and the human eye will remain the last resort.

As for security, the photograph in a document is not the only factor to inform over sincerity or mischief of a passenger: Besides other indicators such as the quality of the document paper, stamps, place of issue etc., reports from security personnel revealed that it is mostly the passenger's behaviour which raises suspicions. A sweating person with insecure behaviour might reveal dishonest intentions much more reliably than any photograph. In this sense, it is custom already at places of high security to look out for suspicious behaviour. According to a recent article in *Der Spiegel* (Blech, 2008) security checks are redesigned in ways that the atmosphere is benevolent, the majority of the passenger feels relaxed, and therefore nervous persons can be identified more easily. Also, the questioning of passengers before boarding an air plane to high risk countries, for example, is a measure to filter out suspicious behaviour. For the security personnel performing such job it is therefore not so much the picture in the travel document, but the facial expression and its authenticity that is critical. Reliable and differentiated abilities in reading a person's emotional state are required to do this job. Part III of this work is designed to look at this issue of emotion recognition more closely.

Part III

Emotion Recognition

9. Recognition of Emotion in Moving and Static Composite Faces

9.1. Introduction

Components and Configurations

Recognizing people's identity and their emotional state is a basic and important skill in social interaction which we perform with great accuracy and consistency. As mentioned in Part I of this work, a common classification of the information contained in faces is the distinction between component information, relating to separable local elements such as eyes, mouth, or nose (e.g. Carey & Diamond, 1977), and configural information, referring to the spatial relations of these elements (e.g. Schwaninger, Wallraven, Cunningham, & Chiller-Glaus, 2006). Several hypotheses have been proposed to explain the mechanisms used in adult face processing, although general consensus holds that for the recognition of faces, configural information is of special importance.

The so called composite effect (Young, Hellawell, & Hay, 1987) is an impressive demonstration of the importance of configural information in face recognition (see Figure 1.1 in Part I): When combining the top half of one face with the bottom half of another face in alignment, recognition is significantly impaired in respect to misaligned halves. This fact is generally explained by the fusing of the aligned halves to one single identity, which supports the view that the components of a face are not the main source upon which we draw when recognizing a face.

Configural information is not only central in the recognition of facial identity, as has been discussed so far, but also in the processing of emotion. McKelvie (1995) could show that emotions which were easily recognized in upright orientation could not be named when inverted. Inversion being known to disrupt configural processing (Yin, 1969), the results of McKelvie thus lead to the assumption that also in emotion processing configural information is of central importance. Similar results were attained by Prkachin (2003). Furthermore, Young, Hellawell, and Hay's composite paradigm, originally designed for facial identity, proved to be also applicable to expression recognition (Calder, Young, Keane, & Deane, 2000).

Dynamic information

Faces are not static objects, but are constantly in motion. As already discussed in Part I, motion could contribute to face recognition by several mechanisms, e.g. by supplemental information due to increased number of views available, by building up a 3D-representation, or by a quality of its own inherent to dynamic information (e.g., Lander & Bruce, 2000, for the recognition of facial identity; O'Toole, Roark, & Abdi, 2002, for face recognition in general). Humans have been found to encode and represent temporal information about expressions (Edwards, 1998). For both the recognition of identity and expression, however, research up to date has not established a clear answer weather motion facilitates recognition or not (O'Toole et al., 2002). Evidence for a beneficiary effect of motion in identity recognition is given for example by Knappmeyer, Thornton and Bülthoff, (2003) or Lander and Bruce (2004, 2000). Also, under suboptimal viewing conditions such as poor illumination or long distance, dynamic information proved to be helpful (Lander, Christie, & Bruce, 1999). There are, however, converse findings: Bruce, Henderson, Greenwood, Hancock, Burton, and Miller (1999) demonstrated difficulties in matching unfamiliar target faces on video against arrays of photographs, where accuracy proved to be poor in the static condition even when viewpoint and facial expression were standardized, and did not improve when the target face was shown in motion. Christie and Bruce (1998) confirmed the lack of improvement. For the recognition of expressions, Bassili (1978) showed that facial expression undergoing dynamic change – seen as white dots on a black-masked face – were perceived correctly even when feature-based information was eliminated, while this was not the case for emotions under static conditions. But in a study manipulating the velocity of change in emotion, Kamachi, Bruce, Mukaida, Gyoba, Yoshikawa, and Akamatsu (2001) found that the overall performance was nevertheless slightly poorer for dynamic images than for static images.

A study conducted by Ambadar, Schooler, & Cohn (2005) addressed the question what aspect exactly in the recognition of emotions was enhanced by motion. They used subtle facial expressions to test the four hypotheses that a beneficial influence of motion might be due to a) a denser sampling of pictures, b) dynamic information, c) the facilitation of configural processing, and d) enhancement of the perception of change. In their study, they used four conditions to assess these questions: in a static condition, they showed only one single static facial expression; in a multi-static condition, they used all slides contained in a short video-sequence from neutral to expressive, albeit separated with a mask to disrupt any dynamic information; in a dynamic condition, they showed a video sequence from neutral to expressive, and in a first-last condition they only showed the first and the last slide of the video sequence. Ambadar et al. found that recognition performance was equal in the conditions static and multi-static, and was significantly better in the dynamic condition. This implies that dynamic presentation is not superior to static presentation due to denser sampling. Also, performance under the first-last condition did not differ from the dynamic condition, therefore ruling out the possibility that dynamic presentation was superior to static presentation due to a type of information inherent to the dynamic impression itself. When viewing the stimuli in inverted orientation, performance was significantly poorer for the inverted condition than the upright condition. Inversion is known to disrupt configural information (e.g. Bartlett & Searcy, 1993; Carey & Diamond, 1977; Diamond & Carey, 1986; Leder & Bruce, 2000; Searcy & Bartlett, 1996; Sergent, 1984). Since no interaction occurred between motion and orientation, Ambadar et al. reasoned that dynamic information did not specifically support the processing of configural information. They concluded that the only explanation for the superiority of dynamic presentation derived from their study was that dynamic information enhanced the perception of change, as could be seen in the equal recognition performance of the two conditions dynamic and first-last.

Research question

The composite face paradigm, as used in the studies by Young et al. (1987) and Calder et al. (2000), provides a good tool to assess the role of configural and component information in expression recognition. For static stimuli, it has been shown that misalignment of the two halves of the face disrupts configural information, therefore preventing the fusion of the two halves to one impression, which results in better recognition performance of the separate halves when compared to alignment. So far, in research on the composite effect only static stimuli have been used. In reality, however, we usually encounter dynamic faces. By using inverted stimuli, Ambadar et al. (2005) could show that dynamic information does not specifically enhance configural information. This being one of the first studies to investigate the role of dynamic information on the processing of components and configurations, it is the aim of Part III to investigate whether the findings of Ambadar et al. are indeed of general nature, or if they are an artifact of one specific method, namely the inversion paradigm. Thus, the role of motion in the processing of configural and part-based information is explored within the framework of the composite face paradigm. The reasoning is that if the effects found by Ambadar et al. withstand the test of being subjected to a different method than the one originally used, the results might indeed be universal. The composite paradigm provides a very potent tool to explicitly assess the role and interaction of configural and component information. If dynamic information enhances configural processing but not part-based processing, an interaction between alignment of the halves and motion can be expected in the direction of a greater difference between the performance on aligned and misaligned halves for dynamic stimuli than static stimuli. On the other hand, if dynamic information enhances part-based processing but not configural processing, such an interaction is expected to point to the other direction, namely to a decreased difference in performance between aligned and misaligned halves for dynamic compared to static stimuli. If no interaction occurres between alignment and motion, the results can be interpreted in ways that dynamic information does not enhance either configural nor component information, but has a quality to itself, be this reduction of change blindness, as suggested by Ambadar et al, or something else. A second aim of this study concernes the recognition performance of different emotions: In a pilot study using stimuli from the Ekman and Friesen series (1976), Calder et al. found that happiness and disgust were better recognized from their top halves, anger, fear, and sadness from their bottom halves, and surprise from both. According to this finding, they designed their following studies on the composite effect. Since the Ekman stimuli only exist as static images and not as video sequences of the developing emotions, a different set of stimuli of which a moving and a static version exists has to be used. As it is unclear to what extent these newly created stimuli are comparable to the Ekman stimuli in terms of their expressiveness in their top and bottom sections, as well as for the sake of completeness, the experiments in Part III therefore involve the combination of all emotions with each other and thus address the question of dominance of one half over the other from a new angle. One last research question concernes the interaction between dynamic information and the dominance of the top and bottom halves: The time course of a developing emotion might lead to otherwise poorly recognizable halves to be more easily recognized if presented in motion rather than statically. Such change of dominance would imply that otherwise inexpressive features of the face gain with dynamic information, and thus support the enhancement of component information by motion. In order to address these questions, two experiments were conducted: Experiment 8a with static composite stimuli, Experiment 8b with moving composite stimuli.

9.2. Experiment 8a: Static Images

The aim of Experiment 8a was to analyze the role of configural and part-based information on the recognition of facial expressions in a composite face paradigm. More specifically, a baseline of recognition performance with static stimuli was to be defined and in a next step be compared to a moving condition (Experiment 8b). Also, Experiment 8a had the purpose to replicate the findings by Calder et al. (2000) on the composite effect. The reason for this comparison was to validate the set of stimuli used in Experiment 8a against the well-researched Ekman and Friesen series (1976) for later use in a moving version (Experiment 8b).

Method

Participants

Twenty-four students (16 female), aged between 20 and 39 years (*M* = 25), of the University of Zurich voluntarily participated in this experiment. All had normal or corrected-to-normal vision.

Materials

Colored freeze images of peak expressions of four actors (all female) expressing the six basic human emotions (anger, fear, surprise, sadness, disgust, happiness) were used as stimuli. The 24 photographs were divided into a top and a bottom segment by cutting along a horizontal line through the bridge of the nose. These halves were recombined in a way that every top emotion was joined with every bottom emotion of the same actor, resulting in a total of 36 stimuli per actor. From these stimuli, two versions were created according to the design used by Young et al. (1987): In the aligned condition (AL), the two halves were combined to form one single face, while in the misaligned condition (ML) the halves were shifted horizontally so that the nose of the top half was in alignment with the edge of the face of the bottom half. For 50% of the misaligned stimuli, the top half was shifted to the right of the bottom half, and for 50% to the left (counterbalanced across emotion and composition). In total, a set of 288 stimuli was created (6 * 6 expressions * 4 actors * 2 conditions, e.g. aligned vs. misaligned). An example of the stimuli is displayed in Figure 9.1.

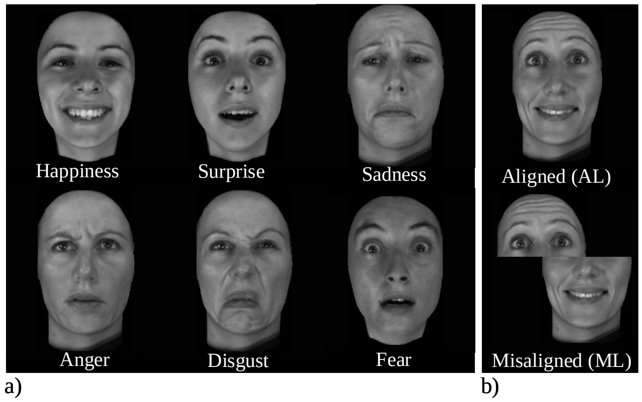


Figure 9.1: Example of the stimuli used in Experiment 8a. a) six basic emotions (different emotions are played by different actors). b) aligned and misaligned composites. The top half is fear, the bottom half happiness.

Design and procedure

Three within-subjects factors were investigated: stimulus type (aligned, misaligned), composition (same emotion, different emotion), and emotion (happy, sad, surprised, angry, disgusted, fearful). The dependent variable was the number of correct answers (recognition of facial expression). The stimuli were presented in the center of a 15” screen with a resolution of 600x800 pixels. The viewing distance of 60 cm was maintained by a head rest, the stimuli covered a vertical visual angle of 6°. The experiment began with a warm-up session of 12 trials, followed by the main session of 288 trials. The factors emotion, composition and alignment were counterbalanced across four blocks; presentation of trials within each block was randomized. The participants’ task was to identify the emotion depicted in each stimulus. A fixation cross (display duration: 1000 ms) preceded each stimulus. Stimuli were presented for maximally 7 seconds (participants could choose to stop the presentation before these 7 seconds if the wanted) and followed by six answer buttons labeled with the names of the six emotions (anger, fear, surprise, sadness, disgust, happiness) to be pressed with the left mouse button. There was no time limit to response. The participants were free to take a short brake after the first, second, and third block of the experiment. Half of the participants were instructed to assess the facial expression depicted in the top segment, the other half to assess the bottom segment. Each session lasted approximately 50 minutes.

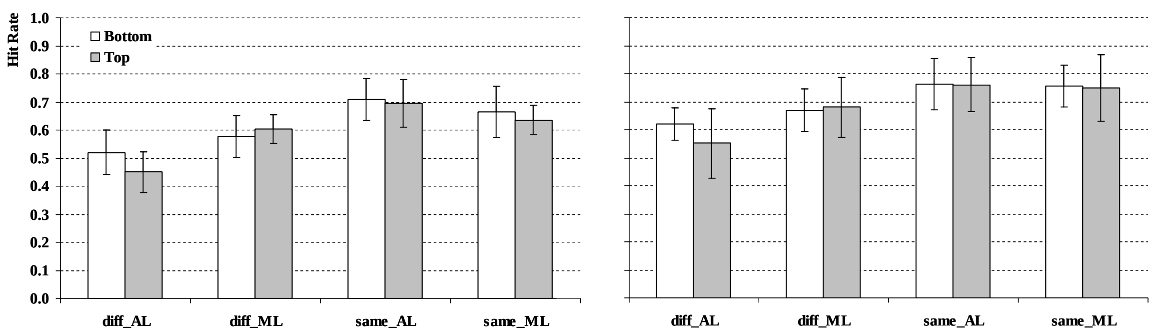
The hypothesis was that if facial expression recognition was largely based on configurations, more errors were expected for aligned composites because the two emotions would fuse to create a new percept which made it difficult to identify either half. If the recognition of facial expression was rather based on the components, no differences for the aligned and misaligned conditions were expected.

Results and Discussion

The principal data analysis involved correct responses (Hits). In the following, responses to the top and bottom halves of the stimuli are analyzed separately. To get an overall impression, in a first step the data of all emotions were analyzed together. In a second step, every emotion was analyzed separately. Results are displayed in Figure 9.2a.

Bottom: A two-way ANOVA with the factors composition (same emotion, different emotion) and alignment (aligned, misaligned) revealed a main effect of composition (F (1, 11) = 31.93, p < .001, η^2 = .75). The interaction between alignment and composition was also significant (F (1, 11) = 12.42, p < .01, η^2 = .53. There was no main effect of alignment (F (1, 11) = 0.12, p = .73, η^2 = .01).

Top: The two-way ANOVA with the factors composition and alignment revealed a main effect of both composition (F (1, 11) = 86.51, p < .001, η^2 = .86) and alignment (F (1, 11) = 29.02, p < .001, η^2 = .73), the interaction between composition and alignment was also significant (F (1, 11) = 31.73, p < .001, η^2 = .74).



a) Static (Experiment 8a) b) Moving (Experiment 8b)

Figure 9.2: Effect of composition and alignment on recognition performance (average of all emotions) for top and bottom halves. a) static, b) moving. diff = different emotion, same = same emotion, AL = aligned, ML = misaligned. Error bars represent standard deviations.

The data for the top half of the face revealed a robust composite effect in emotion recognition. The findings are consistent with the study conducted by Calder et al. and Young et al. The significant effect of alignment on recognition performance can be interpreted as evidence for holistic or configural processing in the sense that aligned halves seemed to fuse into one single expression and therefore reduced recognition of one single half. This interpretation is further sustained by the significant interaction between alignment and composition: While faces depicting the same emotion in both halves were recognized more easily in the aligned condition, the opposite was the case for faces composed of different emotions. Misalignment disturbed recognition performance of same-emotion faces, since the configural processing of the face as a whole was disrupted. On the other hand, the same disruption was beneficial for different-emotion faces by preventing the fusion of both halves into a one single impression. Results from the bottom half are somewhat different in the sense that no significant effect of alignment was found. This implies that the recognition of the bottom half not only relies on configural information, but includes the processing of separate parts as well. The interaction between the alignment and composition, nevertheless, stresses the prevailing influence of configural information.

One has to bear in mind, however, that the findings described so far are an average of all six emotions tested. In a three-way ANOVA with the factors emotion (anger, disgust, fear, happiness, sadness, and surprise), composition and alignment, a main effect of the factor emotion was found for both top and bottom halves (Bottom: F (5, 55) = 28.40, $p < .001$, η^2 = .72; Top: F (2.97, 32.66) = 30.40, $p < .001$, $\eta^2 = .73$). This implies that the recognition pattern was not identical for all expressions, and therefore they have to be interpreted separately. Results are displayed in Figure 9.3.

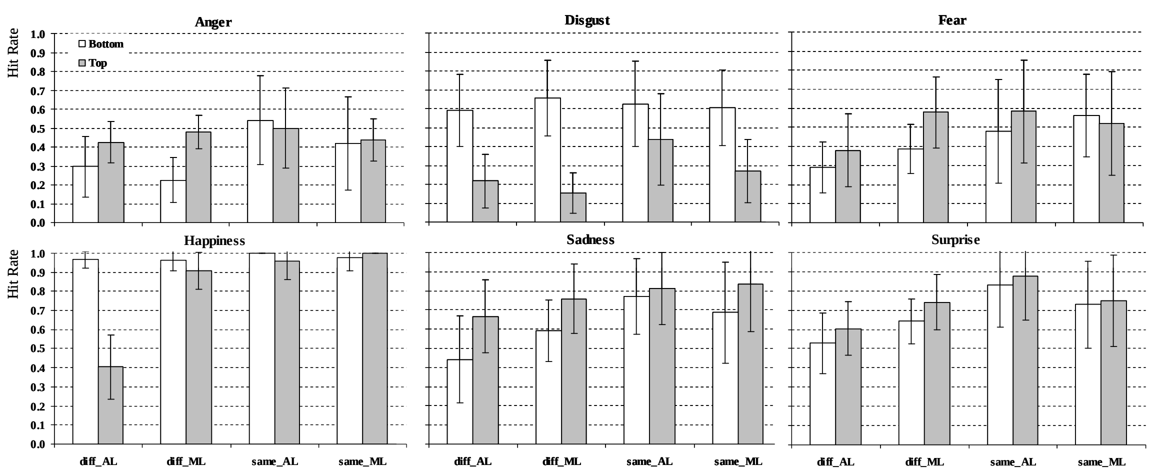


Figure 9.3: Static (Experiment 8a): Effect of alignment and composition on recognition performance for top and bottom halves, for each emotion. Error bars represent standard deviations. diff = different emotion, same = same emotion, AL = aligned, ML = misaligned.

As can be seen in Figure 9.3, the separate emotions lead to a variety of different recognition patterns, not all of them consistent with the above described average of all emotions. One central question of this study was whether the processing of facial expressions was mainly based on configurations, and to what extent also part-based processing played a role. In this design, the crucial test to this question was whether a significant interaction between composition and alignment occurred: a significant interaction would indicate that alignment affected same-emotion faces and different-emotion faces in unequal ways, in particular that different-emotion faces were more difficult to recognize when aligned than when misaligned, while for same-emotion faces the opposite was the case. Such an interaction could be interpreted as evidence for the fusion of the two separate halves of the face, beneficial in the case of same-emotion faces and detrimental for different-emotion faces, which would stress the pivotal role of configural information in expression recognition. Also, a main effect of composition would furthermore corroborate the integration of the two halves of the face and therefore the dominance of configural information. On the other hand, if mainly part-based information was processed in expression recognition, there would be no influence of either composition or alignment on recognition performance, and thus neither a significant interaction between composition and alignment, nor a main effect of each, would be expected. Results from two-way ANOVAs with the factors composition and alignment on every separate emotion are displayed in Table 9.1 (left side).

		Static (Experiment 8a)			Moving (Experiment 8b)			Influence of Motion			
		Align	Comp	Align* Comp	Align	Comp	Align* Comp	Main effect	Motion* Align	Motion* Comp	Motion* Align*Comp
Bot	Anger	.07	***	.49	*	.57	.47	**	.62	*	.89
	Disgust	.66	.81	.25	.86	.56	.61	*	.81	.55	.21
	Fear	*	**	.88	.09	.08	.58	.12	.97	.36	.76
	Happiness	.46	.05	.39	*	**	*	.39	.08	.07	.31
	Sadness	.45	**	**	*	***	.21	.19	.38	.39	.32
	Surprise	.92	*	**	1.00	**	.17	.25	.94	.80	.14
Top	Anger	.90	.64	.54	.75	.47	.66	.43	.73	.79	.43
	Disgust	*	**	.33	.23	*	.45	***	.27	.74	.75
	Fear	.11	.28	**	.65	.33	.13	.09	.60	.95	.54
	Happiness	***	***	***	***	***	***	.64	.61	.49	.78
	Sadness	.23	*	.39	.07	.06	.37	.45	.91	.98	.75
	Surprise	.88	**	*	*	**	.80	.88	.34	.42	.07

Table 9.1

Left: Effect of composition and alignment on recognition performance for the static condition (Experiment 8a): p-values of a two-way ANOVA. *Middle:* Effect of composition and alignment on recognition performance for the moving condition (Experiment 8b): p-values of a two-way ANOVA. *Right:* Effect of motion on recognition performance of both the static and moving condition: p-values of a three-way ANOVA. NOTE: * = $p < .05$, ** = $p < .01$, *** = $p < .001$. Error bars represent standard deviations.

According to this criteria, none of the six tested emotions qualified for purely configural processing. The fact that for most of them at least one of the expected effects was not significant indicates that, to a certain extent, recognition of facial expression also draws upon part-based information. Only the processing of individual features without involving the whole facial context could account for equal recognition performance for aligned and misaligned composites. Accordingly, different-emotion and same-emotion composites could only be equally recognized under the assumption that the two halves did not fuse to a single expression, therefore involving also feature information. Thus, these data suggest that for the processing of facial expressions, both configural and part-based information was assessed, the extent of involvement of each type of information depending on the specific emotion. Three cases, namely the top half of anger and the bottom half of both disgust and happiness even seemed to solely rely on part-based information, since no difference in recognition performance was found between the aligned and misaligned condition, and between the same-emotion and different-emotion condition. In the case of the bottom half of happiness, however, recognition performance was close to ceiling, which does not allow a reliable interpretation of the data. Neither does the top half of happiness, since the conditions same_ML was at ceiling (overall performance = 1.0).

A second aim of this study was to compare the stimuli of Experiment 8a to those created by Ekman in order to assess the former's validity. Using stimuli of the Ekman series, Calder et al. assessed for each of the six emotions whether it was better recognized in the top or in the bottom half, or in different terms, which half is dominant in any specific emotion. If similar patterns of dominance as described by Calder et al. could be found with the stimuli of Experiment 8a, the two sets of stimuli would be comparable in terms of their expressiveness. As can be derived from Figure 9.3, the top and bottom halves of the face evoked different recognition patterns. For each emotion, the dominant half was determined using an independent-sample t-Test on the overall recognition performance of top and bottom halves. The dominance along with the p-values of the t-Test are displayed in Table 9.2 (left side).

	Static (Exp. 1)		Moving (Exp. 2)		Change of Dominance
	Dominance	p-value t-test	Dominance	p-value t-test	
Anger	Top	**	Bottom	.64	No
Disgust	Bottom	***	Bottom	**	No
Fear	Top	.06	Top	*	No
Happiness	Bottom	***	Bottom	**	No
Sadness	Top	*	Top	*	No
Surprise	Top	.12	Bottom	.75	No

Table 9.2

Dominance of the top or bottom half of the faces. *Left:* static condition (Experiment 8a), *middle:* moving condition (Experiment 8b), *right:* Change of dominance between moving and static. NOTE: * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

Calder et al. reported that in their study, anger, fear, and sadness were better recognized from the top half, disgust and happiness from the bottom half, and surprise from both. The results of Experiment 8a are highly consistent with these reports. Interpreting a non-significant difference between the two halves as the emotion being equally recognized from both top and bottom halves, it was also found that anger and sadness were better recognized from the top half, disgust and happiness from the bottom half, and surprise from both. With the sole exception of fear, where according to Calder et al. the top half was dominant while here a significant difference was missed narrowly, all other findings were identical. Thus, the conclusion could be drawn that the stimuli used in Experiment 8a were consistent in terms of expressiveness with the Ekman stimuli and could therefore be used in a moving version for Experiment 8b.

9.3. Experiment 8b: Moving Images

Experiment 8a was conducted with moving stimuli in order to investigate the influence of dynamic information on the processing of facial expressions.

Method

Participants

Twenty-four students (13 female), aged between 22 and 39 years (M = 26) of the University of Zurich voluntarily participated in this experiment. All had normal or corrected-to-normal vision.

Materials

In order to analyze the effect of motion on the processing of facial expression, seven-second video sequences of facial emotions (happiness, fear, disgust, surprise, anger, sadness) ranging from neutral expression to maximal emotion and back to neutral were recorded from the same four female actors. Using computer simulation, the video sequences were edited such as to show colored faces on black background. Along the lines of Experiment 8a, aligned and misaligned versions were created of the video sequences, resulting in a total of 288 stimuli.

Design and Procedure

The same factors as in Experiment 8a were investigated. The procedure was as described in Experiment 8a. Display duration of the stimuli was always 7 seconds.

Results and Discussion

The principal data analysis involved correct responses. As in Experiment 8a, responses to the top and bottom halves of the stimuli were analyzed separately.

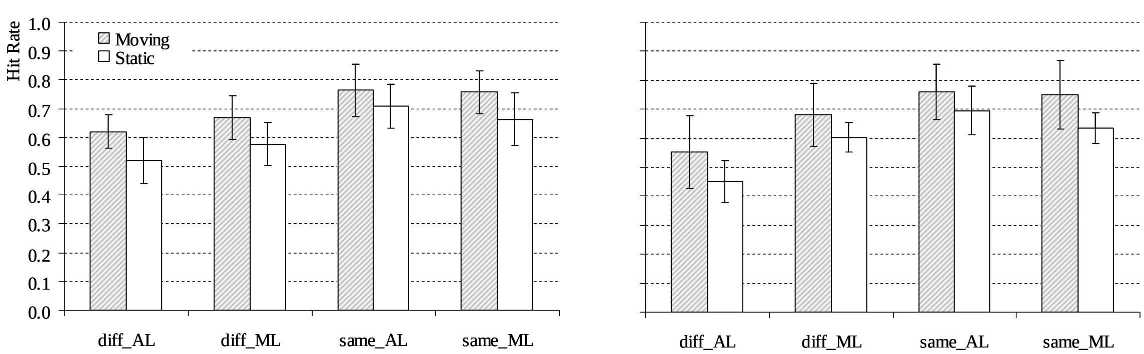
Bottom: A two-way ANOVA with the factors composition (same emotion, different emotion) and alignment (aligned, misaligned) revealed a significant main effect of alignment (F (1, 11) = 31.26, $p < .001$, $\eta^2 = .74$) and a significant interaction between the two factors (F (1, 11) = 6.37, $p < .05$, $\eta^2 = .37$). The factor alignment did not lead to a significant main effect (F (1, 11) = 4.05, $p = .07$, $\eta^2 = .27$). However, η^2 being reasonably high, the lack of a significant effect might be due to the small number of participants. Otherwise, results reflect the findings of the static condition in Experiment 8a.

Top: The same two-way ANOVA revealed a significant main effect of both composition (F (1, 11) = 32.43, $p < .001$, $\eta^2 = .75$) and alignment (F (1, 11) = 18.80, $p < .01$, $\eta^2 = .63$), as well as a significant interaction between the two factors (F (1, 11) = 9.88, $p < .01$, $\eta^2 = .47$). Results are displayed in Figure 9.2b.

These data are highly consistent with the results obtained in Experiment 8a and corroborate the position that the processing of facial expression strongly relies on configural information, for both static and moving stimuli. It is not clear, however, to what extent motion enhances or reduces recognition performance when compared to static stimuli. For this purpose, the static condition was directly compared to the moving condition, which allowed investigating to what extent non-rigid motion affected the perception of facial emotion and whether this effect was different for configural versus part-based processing. The data for the static condition was adopted from Experiment 8a. Results are displayed in Figure 9.4.

Bottom: Correct responses were submitted to a three-factor ANOVA with the factors motion (moving, static), composition (same emotion, different emotion) and alignment (aligned, misaligned). The ANOVA revealed a significant main effect for the factors composition (F (1, 22) = 62.81, $p < .001$, $\eta^2 = .74$) and motion (F (1, 22) = 12.79, $p < .01$, $\eta^2 = .37$). The interaction between composition and alignment was significant (F (1, 22) = 18.78, $p < .001$, $\eta^2 = .46$). Alignment did not reveal a significant effect (F (1, 22) = 1.72, $p = .20$, $\eta^2 = .07$). No significant interactions between motion and other factors were found.

Top: The same three-way ANOVA as conducted for the bottom half revealed a significant main effect for all three factors: composition: F (1, 22) = 87.89, $p < .001$, $\eta^2 = .80$, alignment: F (1, 22) = 43.22, $p < .001$, $\eta^2 = .66$, and motion: F (1, 22) = 8.73, $p < .01$, $\eta^2 = .28$. The interaction between composition and alignment was significant, F (1, 22) = 36.70, $p < .001$, $\eta^2 = .63$. Again, no significant interactions between motion and other factors were found.



a) Bottom b) Top

Figure 9.4: Recognition performance for both static (Experiment 8a) and moving (Experiment 8b) condition, average of all emotions. a) bottom half, b) top half. Error bars represent standard deviations. diff = different emotion, same = same emotion, AL = aligned, ML = misaligned.

For both the top and the bottom halves, in all four conditions, recognition was better for moving stimuli than for static stimuli, as is reflected in the significant main effect of the factor motion. These results are consistent with other findings on the beneficial influence of motion on expression recognition (see Introduction of Part III). One aim of this study was to investigate whether such a beneficial influence was mediated by enhanced processing of configural or part-based information. The lack of any significant interaction between motion and other factors suggests that it is neither exclusively configural nor only part-based processing that is boosted by motion; rather it seems that motion supports configural and part-based information alike. These data, however, represent an average of all emotions, and Experiment 8a revealed that not all emotions are processed alike. The emotions therefore have to be analyzed separately in order to reliably decide on the influence of motion on expression recognition. Results of this analysis are displayed in Figure 9.5.

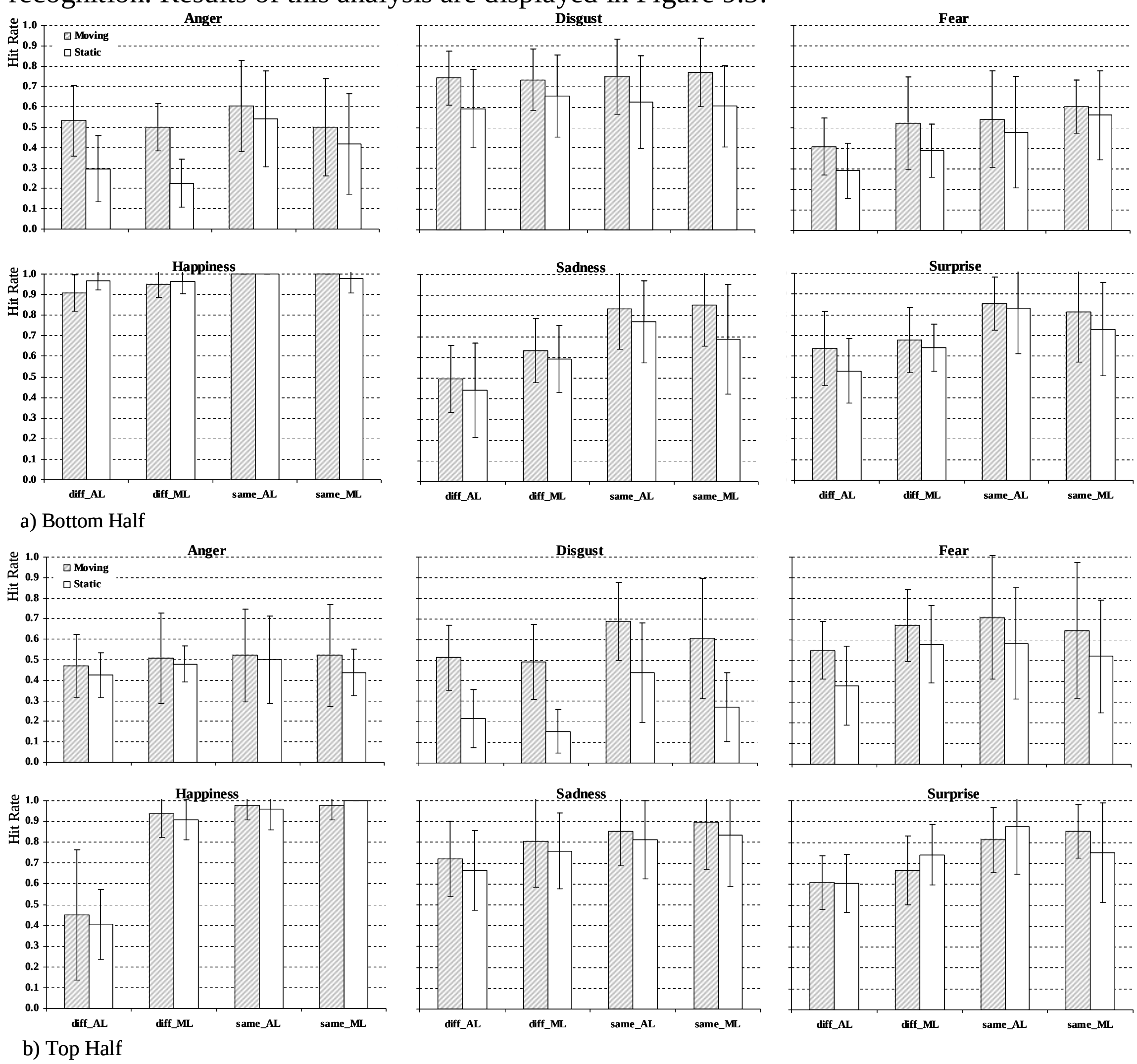


Figure 9.5: Recognition performance for both static (Experiment 8a) and moving (Experiment 8b) condition, for every emotion separately. a) Bottom half, b) top half. Error bars represent standard deviations. diff = different emotion, same = same emotion, AL = aligned, ML = misaligned.

As with the average of all emotions, also for separate emotions in most cases the moving condition lead to increased recognition performance compared to the static condition. On the lines of Experiment 8a, for every emotion, a two-way ANOVA with the factors composition and alignment was conducted in order to investigate whether in the moving condition the processing of one specific emotion relied mainly on configural or on part-based information. Results are displayed in Table 9.1 (middle). As with the static stimuli, none of the emotions seemed to be processed purely on the basis of configural information; the performance of happiness – the only emotion with significant results in all three analyses – was close to ceiling, which does not allow an interpretation of the data. To compare the results of Experiment 8a and 8b, a three-way ANOVA with the factors motion (static, moving), composition and alignment was conducted. Results are displayed in Table 9.1 (right).

While in the analysis on all emotions together (see Figure 9.4) the factor motion was significant for both top and bottom halves, in the separate-emotion analysis only disgust (top and bottom) and anger (bottom) showed a significant main effect; for all the other emotions, motion did not significantly increase recognition performance. This difference between the all-emotion analysis and the separate-emotion analysis might arise from the smaller number of cases in the latter (288 trials in the all-emotion analysis, 48 in the separate-emotion analysis). Apart from that, result challenges previous reports of a beneficial influence of motion on the recognition of facial expressions. However, there are other studies which, like in this experiment, lacked to find such beneficial influence (e.g. Kamachi et al., 2001). As for the question whether motion specifically augments configural or part-based processing, a significant interaction between motion and the other factors would indicate facilitation of configural processing, while the lack of such a significant interaction would stand for no specific augmentation of either of the two types of information. As can be derived from Table 9.1 (right), for all emotions, of all the interactions between motion and the two other factors, but for one single exception (anger bottom), none was significant. This implies, as was already mentioned in the context of the all-emotion analysis, that motion does support configural and part-based processing in a similar way. Or, as Ambadar et al. (2005) phrased it, the lack of a significant interaction between motion and the other factors “indicates that motion does not improve perception of facial expressions by facilitating configural processing” (p. 406).

Despite the fact that motion does not seem to augment either quality of information specifically, motion still generally renders expressions more easily recognizable. In Experiment 8a, it was shown that every emotion has its dominant half by which recognition is easiest. If motion raises the overall recognition performance, it might well be that the dominance of the halves is influenced by that factor and might even change from top to bottom or vice versa when compared to static stimuli. Analogue to Experiment 8a, an independent-sample t-test was conducted on the overall recognition performance of top and bottom halves. The dominance along with the p-values of the t-Test are displayed in Table 9.2 (middle) to address this question. Even though for anger and surprise the tendency of the dominance switched from top to bottom when compared to the static condition, these effects are not significant and do therefore not represent any real change of dominance. For all the other emotions, the dominance stayed the same, the only difference being that for fear, the tendency to be better recognized from the top half became significant in the moving condition while it was slightly below significance level in the static condition; the direction, in any case, remained the same. From these data it can be inferred that motion does not lead to any change of dominance in emotion recognition.

9.4. General Discussion

In two experiments, the composite paradigm by Young et al. (1987) was used to investigate the effect of motion on expression recognition. For both the top and bottom segments of the faces, a beneficiary effect of motion on the recognition of facial expression was found when taking all emotions together. For separate emotions, however, this significant effect mostly vanished, although a general tendency towards increased performance due to motion remained. These results reflect the non-conclusive findings of other researchers on the nature of dynamic information. Also, motion seems to have different effect on separate emotions: Anger and disgust, which are generally recognized less accurately than other expressions, benefit from motion, while the other expressions do not. This finding might be interpreted along the lines of Lander et al. (1999) according to whom motion positively influences recognition performance under difficult viewing conditions. In this sense, it seems reasonable to argue that expressions which are most difficult to recognize under static conditions make the most of additional dynamic information. Furthermore, despite the lack of any significant influence of motion on recognition performance of separate emotions, it was nevertheless found that motion did at least not reduce recognition performance. Thus, the tendency of these results point in the same direction as the study by Ambadar et al. (2005) who found that all emotions except happiness benefit from motion. In their experiment, they used subtle facial expressions rather than peak expressions. Ambadar et al. argue that the absence of any significant influence by motion might arise from the usage of stimuli: peak expressions, as commonly used in experiments on emotion recognition, might be too intense as not to mask the more subtle effects of dynamic information. This explanation might well account for the missing of a significant effect of motion on separate expression.

For both static and moving stimuli, a robust composite effect was found for top halves, i.e. misaligned faces of different emotions were better recognized than aligned faces, while for faces depicting the same emotions in their top and bottom halves the opposite was the case. For bottom halves, the composite effect proved to be less pronounced, as neither the static not the moving condition produced a significant main effect of alignment. It might be derived from this that components in the lower part of a face are processed more independently from their surroundings than are components in the upper part. The significant interaction between composition and alignment in both halves and both conditions, however, suggests that the processing of facial expressions is based on holistic processes, which reflects the general consensus in face recognition research of configural information playing a pivotal role. The fact that the data analysis of Experiments 8a and b involving correct responses matched the findings by Calder et al. (2000) who focused on reaction times implies that the composite effect is highly robust. Furthermore, the question was addressed whether dynamic information specifically enhanced configural or part-based processing. The composite design allowed making accurate predictions about both types of information independently of one another. Using the inversion paradigm, Ambadar et al. (2005) already suggested that configural processing is not specifically supported by motion. This result was completed in Experiment 8b by the finding that neither was part-based information: The lack of any significant interaction between motion and other factors (alignment and composition) suggests that motion influences the processing of configural and part-based processing alike and does not enhance either of the two processes sepatately. Therefore, it seems likely to assume that motion either has a beneficial effect on recognition performance due to the dynamic information per se, or influences other processes in the recognition of faces, such as change blindness, as suggested by Ambadar et al.

In accordance with Calder et al. (2000), this study that anger and sadness were recognized from the upper half, happiness and disgust from the bottom half, and surprise and fear from both (Calder et al. found that fear was recognized from the top). The assumption that motion might lead to a change of dominance due to enhancement of specific otherwise non-recognizable features could not be supported. This finding thus can be taken as second evidence that motion does not raise components or configurations specifically, but enhances other processes in expression recognition.

A model which integrates the assumptions of both holistic and componential processing was proposed by Schwaninger et al. (2002; 2003). It complements the purely holistic view with empirical findings that facial components and configural information are encoded and stored explicitly when faces are upright. The model has also been implemented using computational modeling and simulations have shown a striking similarity between human and computer model data (Wallraven, Bülthoff, & Schwaninger, 2005). As was shown with this study, the same process seems to apply to moving stimuli. The model by Schwaninger et al. could thus be extended by the factor motion.

Closing Words

This work addressed the issue of face recognition form several angles. In Part I a solid overview was given over the state-of-the-art literature on the topic and the basic concepts of face recognition.

The recognition of a face's emotional expression, which is a highly relevant task in everyday social communication, was addressed in Part III. Since natural face recognition occurs with living and moving persons, the special influence of dynamic information on emotion recognition was assessed. The two experiments revealed that motion indeed has a quality of its own which helps us to correctly determine a person's emotions. But astonishing above that was the fact that the difference between moving and static images was not so large after all. It therefore seems that even from static images, which in fact represent a highly unnatural stimulus to the human eye, reliable information can be drawn. This phenomenon once more discloses our exceptional abilities to recognize a human face under a tremendous variety of circumstances.

Part II focused on the question of how we can identify a person from a photograph, as is the task of a great number of security personnel in a large variety of occupational fields, such as border controllers, police officers, cashiers, bank clerks, train conductors etc. It was shown that the human ability to perform this task was very limited indeed. This phenomenon, although very plausible to those who have ever tried serious identification from a photograph, is nevertheless a striking one when considering that the current position in the scientific community holds that we are great experts in face recognition. Furthermore, it is irritating to think that at thousands of border crossings, credit card transactions, eyewitness testimonies etc. worldwide a task is required from security personnel which actually very clearly overtaxes their cognitive abilities.

The conclusions for application drawn from this research are threefold: Evidently, photographs in documents do not provide ground enough for sufficient identification. One option would therefore be to raise the ability for identification trough training. In the field of X-Ray screening such training systems have already been successfully implemented (e.g. Koller et al., (2008). The studies presented in this work indicate that training might indeed raise identification performance to a certain degree; whether this increase is sufficient and can be reached within useful time, however, is open to debate and would require substantial efforts on the development of a more sophisticated training system, in particular with a much larger stimulus library. Should training turn out to be inapplicable, a second option would be the use of automated face recognition for identification from photographs. As mentioned earlier, such machines are already operational at several airports[13]. They are, however, still not accurate enough with photographs of lesser quality and cannot yet be implemented on a larger scale. Whether computers will one day replace the human eye at sites of identification remains to be seen over the next years. A third option, probably the most likely one, is to simply accept that the human face – subject to most drastic changes through hairstyle, external paraphernalia, beards, or the passing of time – might just not be the best indicator of identity, and therefore to introduce other measures of identification, such as biometrical data, as is already done in certain passports. Also, the lookout for suspicious behavior at border crossings to identify passengers with ill intentions can be regarded as such an alternative measure. However, at points of lesser relevance than international border crossings, such as everyday interactions where identification is required in trains, offices, purchases etc., the situation is probably likely to remain as it is. After all, a world of absolute control and surveillance is not realistic in the first place, and – to most of us – not desirable.

[13] Neue Zürcher Zeitung, Mittwoch, 9. Juli 2008, p. B1

References

Books

Echt falsch. Leitfaden zur Erkennung von Fälschungen. (2006). Kriminalteschnische Abteilung der Kantonspolizei Zürich. Swiss Police Edition, kik AG.

Bruce, V. (1988). *Recognising faces.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Buckhout, R., & Regan, S. (1988). Explorations in research on the other-race effect in face recognition. In: M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Volume 1. Memory in everyday life* (pp. 40-46). New York: Wiley.

DeRenzi, E. (1997). Prosopagnosia. In: T. E. Feinberg & M. J. Farah (Eds.), *Behavioral neurology and neuropsychology* (pp. 245-256). New York, McGraw-Hill.

Ekman, P. & Friesen, W. V. (1976). *Pictures of facial affect.* Palo Alto, California: Consulting Psychologists Press.

Ekman, P. & Friesen, W. V. (1978). *Facial action coding system.* Palo Alto, Consulting Psychologists Press.

Ellis, A. W. (1992). Cognitive mechanisms of face processing. In: V. Bruce, A. Covey, A. W. Ellis, & D. P. Perrett (Eds.), *Processing the facial image* (pp113-119). New York, NY: Clarendon Press.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston, MA: Houghton Mifflin.

Goldstein, A. G. & Chance, J. E. (1981). Laboratory Studies in face recognition. In: G. M. Davies, H. D. Ellis, & J. W. Sheperd (Eds.), *Perceiving and remembering faces,* London, England: Academic Press.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Johnston, R. A., & Ellis, H. D. (1995). The development of face recognition. In: T. Valentine (Ed.), *Cognitive and computational aspects of face recognition, explorations in face space* (pp. 1-23). London, New York: Routledge.

Leventhal, H. & Sharp E. (1965). Facial expression as indicators of distress. In Tomkins S. S. & Izard C. E. (Eds.), *Affect, cognition and personality: empirical studies* (pp. 296-318). Springer, New York.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection Theory: A user's guide.* Cambridge: Cambridge University Press.

Nummenmma, T. (1964). *The language of the face.* Jyvaskyla studies in education, psychology, and social research. Jyvaskyla, Finland.

Plutchik, R. (1962). *The emotions: facts, theories, and a new model*. New York: Random House.

Rock, I. (1973). *Orientation and form.* New York: Academic Press.

Schwaninger, A., Carbon, C. C., & Leder, H. (2003). Expert face processing: specialization and constraints. In: G. Schwarzer & H. Leder, *Development of face processing* (pp. 81-97), Göttingen: Hogrefe.

Schwarzer, G., & Leder, H. (2003). *The development of face processing.* Göttingen: Hogrefe.

Journal articles

Adolphs, R. (2002) Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews, 1(1),* 21-61.

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the Enigmatic Face. The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science, 16(5*), 403-410.

Andrews, T. J. & Ewbank, M. P. (2004) Distinct representations for facial identity and changeable aspects of faces in human temporal lobe. *Neuroimage, 23,* 905-913.

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General, 104,* 54-75.

Bartlett, J. C. & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25,* 281-316.

Barton, J. J. S., Deepak, S., & Malik, N. (2003). Attending to faces: Change Detection, Familiarization, and Inversion Effects. *Perception, 32,* 15-28.

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology: Human Perception and Performance, 4(3),* 373-379.

Bassili, J. N. (1979). Emotion recognition: The role of facial motion and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology, 37,* 2049-2059.

Bentin, S., Deouell, L. Y. & Soroker, N. (1999). Selective visual streaming in face recognition: evidence from developmental prosopagnosia. *NeuroReport, 10,* 823–827.

Blech, J. (2008). Striptease vorm Abheben. *Der Spiegel, 26,* 139.

Bobes, M. A., Martín, M.. Olivares, E. & Valdés-Sosa, M. (2000). Different scalp topography of brain potentials related to expression and identity matching of faces. *Cognitive Brain Research, 9,* 249-260.

Boutet, I., Collin, C., & Faubert, J. (2003). Configural face encoding and special frequency information. *Perception and Psychophysics, 65(7),* 1087-1093.

Bruce, V. & Young, A. W. (1986). Understanding face recognition. *British Journal of psychology, 77,* 305-327.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of Face Identities From Images Captured on Video. *Journal of Experimental Psychology: Applied, 5(4),* 339-360.

Bruce, V., Henderson, Z., Newman, C., & Burton, M.A. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied, 7, 207-218.*

Bukach, C. M, Gauthier, I., & Tarr M. J. (2006). Beyond faces and modularity: the power of an expertise framework. *Trends in Cognitive Sciences. 10,* 159-166

Burton, M. A., Wilson, S., Cowan M., & Bruce, V. (1999). Face recognition in poor quality video: evidence from security surveillance. *Psychological Science, 10(3),* 243-248.

Calder, A. J. and Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience, 6,* 641-651.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology, 26(2),* 527-551.

Carey, S. & Diamond, R. (1977). From piecemeal to configural representation of faces. *Science, 195,* 312-314.

Carey, S. (1992). Becoming a face expert. *philosophical transactions of the royal society of London, 335,* 95-103.

Chance, J. E. & Goldstein, A. G., & McBride, L. (1975). Differential experience and recognition memory of faces. *Journal of social psychology, 97,* 243-253).

Chiroro, P. & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *Quarterly journal of experimental psychology: Human experimental psychology, 48(A),* 879-894.

Christie, F. & Bruce, V. (1998). The role of dynamic information in the recognition of unfamiliar faces. *Memory and Cognition, 26(4),* 780-790.

Collishaw, S. M., & Hole, G. J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception, 29,* 893-910.

Cunningham, D., Kleiner, M., Wallraven, C. & Bülthoff, H. (2005). Manipulating video sequences to determine the components of conversational facial expressions. *ACM Transactions on Applied Perception, 2(3),* 251-269.

Damasio, A. R., Damasio, H. & Van Hoesen, G. W. (1982). Prosopagnosia: anatomic bases and behavioral mechanisms. *Neurology, 32,* 331-341.

Damasio, A. R., Tranel, D. & Damasio, H. (1990). Face agnosia and the neural substrates of memory. *Annual Review of Neuroscience, 13,* 89-109.

Davidoff, J. & Landis, T. (1990). Recognition of unfamiliar faces in prosopagnosia. *Neuropsychologia, 28,* 1143-1161.

Davidoff, J., & Donnelly, N. (1990). Object superiority: A comparison of complete and part probes. *Acta Psychologica, 73,* 225-243.

DeGelder B., Frissen I., Barton J., & Hadjikhani N. (2003). A modulatory role for facial expressions in prosopagnosia. *Proceedings of the National Academy of Science U.S.A., 100(22),* 13105-13110.

Diamond, R. & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115(2),* 107-117.

Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral Cortex, 16(10),* 1453-1461

Doyle, J. M. (2001). Discounting the error costs – Cross-racial false alarms in the culture of contemporal criminal justice. *Psychology public policy and law, 7(1),* 253-262.

Edwards, K. (1998). The face of time: Temporal cues in facial expressions of emotion. *Psychological Science, 9,* 270-276.

Eimer, M. & Holmes, A. (2002). An ERP study on the time course of emotional face processing. *Cognitive Neuroscience and Neuropsychology, 13(4),* 427-431.

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior. 6,* 238-252.

Ellis, H. (1975). Recognizing faces. *British journal of psychology, 66,* 409-426.

Ellison, J. W. & Massaro, D. W. (1997). Featural evaluation, integration, and judgment of facial affect. *Journal of Experimental Psychology: Human Perception and Performance, 23(1),* 213-226.

Farah, M. J., Drain, H. M., & Tanaka, J. W. (1998). What is “special” about face perception? *Psychological Review,* 105 (3), 482-498.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What Causes the Face Inversion Effect? *Journal of Experimental Psychology: Human Perception and Performance, 21(3),* 628-634.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is “special” about face perception? *Psychological Review, 105(3),* 482-498.

Ferguson, D. P., Rhodes, G., Lee, K. & Sriram, N. (2001). “They all look alike to me”: Prejudice and cross-race face recognition. *British journal of psychology, 92(4),* 567-577.

Frijda, N. H., & Philipszoon, E. (1963). Dimensions of recognition of emotion. *Journal of Abnormal and Social Psychology, 66,* 45-51.

Gauthier, I., Behrmann, M. & Tarr, M. J. (1999a). Can face recognition be dissociated from object recognition? *Journal of Cognitive Neuroscience, 11*, 349-370.

Gauthier, I., Skudlarski, P., Gore, J. C. & Anderson, A. W. (2000b). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3,* 191-197.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P. & Gore, J. C. (1999b). Activation of the middle fusiform area increases with expertise in recognizing novel objects. *Nature Neuroscience, 6,* 568-573.

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C. & Anderson, A. W. (2000a). The fusiform „face area” is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience, 12(3),* 495-504.

Gibson, J. J. (1957). Optical motions and transformations as stimuli for visual perception. *Psychological Review, 64,* 228-295.

Goldstein, A. G. & Chance, J. E. (1980). Memory of faces and schema theory, *Journal of psychology, 105(1),* 47-59.

Hanawalt, N, (1944). The role of the upper and lower parts of the face as the basis for judging facial expressions: II. In: Posed expressions and “candid camera” pictures. *Journal of General Psychology, 31,* 23-36.

Hancock, P. J. B., Bruce, V., & Burton, M. A. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Science, 4(9),* 330-337.

Hasselmo, M. E., Rolls, E. T. & Baylis, C. G. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Experimental Brain Research, 32,* 203-218.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L. and Pietrini P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex, *Science, 293,* 2425-2430.

Haxby, J. V., Hoffman, E. A., Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4(6),* 223-233.

Hayward, W. G., Rhodes, G., & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition. *Cognition, 106,* 1017-1027.

Henderson, Z., Bruce, V., & Burton, A.M. (2001). Matching the Faces of Robbers Captured on Video. *Applied Cognitive* Psychology, 15, 445-464.

Henke, K., Schweinberger, S. R., Grigo, A., Klos, T. and Sommer, W. (1998). Specificity of face recognition: recognition of exemplars of non-face objects in prosopagnosia. *Cortex, 34(2),* 289-296.

Hoffman, E. and Haxby, J. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience, 3,* 80-84.

Ingvalson, E. M., & Wenger, M. J. (2005). A Strong Test of the Dual-Mode Hypothesis. *Perception and Psychophysics. 67(1),* 14-35.

Ishai, A., Haxby, J. V., & Ungerleider, L. G. (2002). Visual imagery of famous faces: effects of memory and attention revealed by fMRI. *NeuroImage, 17,* 1729-1741.

Ishai, A., Ungerleider, L. G., Martin, A., & Haxby, J.V. (2000). The representation of objects in the human occipital and temporal cortex. *Journal of Cognitive Neuroscience, 12,* 35-51.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway, *Proceedings of the National Academy of Science, U.S.A., 96,* 9379-9384.

Izard, C. E. (1979). The maximally discriminative facial movement coding system (MAX). Unpublished manuscript. (Available from Instructional Resource Center, University of Delaware. Newark. DE)

Johannson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics, 14,* 201-211.

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception, 30,* 875-887.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception, *Journal of Neuroscience, 17,* 4302–4311.

Kanwisher, N., Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London B*. 361, 2109-28

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: photographs, credit cards and fraud. *Applied cognitive psychology, 11,* 211-222.

Knappmeyer, B., Thornton, I. M., & Bülthoff, H. H. (2003). The use of facial motion and facial form during the processing of identity. *Vision Research, 43,* 1921-1936.

Koller, S. M., Hardmeier, D., Michel, S., & Schwaninger, A. (2008). Investigating training, transfer, and viewpoint effects resulting from recurrent CBT of x-ray image interpretation. *Journal of Transportation Security, 1(2),* 81-106.

Lander, K. & Bruce, V. (2000). Recognizing famous faces: Exploring the benefits of facial motion. *Ecological Psychology, 12(4),* 259-272.

Lander, K. & Bruce, V. (2004). Repetition priming from moving faces. *Memory & Cognition, 32(4),* 640-647.

Lander, K., Christie, F., & Bruce, V. (1999). The role of movement in the recognition of famous faces. *Memory and Cognition, 27(6),* 974-985.

Leder, H. & Bruce, V. (1998). Local and relational aspects of face distinctiveness. *The Quarterly Journal of Experimental Psychology, 51A(3),* 449-473.

Leder, H. & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *The Quarterly Journal of Experimental Psychology, 53A(2),* 513-536.

Leder, H., Candrian, G., Huber, O., & Bruce, V. (2001). Configural features in the context of upright and inverted faces. *Perception, 30,* 73-83.

Liu, C. H., Burton, M. A., Seetzen, H., & Chaughuri, A. (2003). Face recognition is robust with incongruent image resolution: Relationship to security video images. *Journal of Experimental Psychology: Applied, 9(1),* 33-41.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nature Neuroscience, 5,* 910–916.

Malpass, R. S. & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personal and Social Psychology, 1,* 75-86.

McKelvie (1995). Emotional expression in upside-down faces: Evidence for configural and componential processing. *British journal of social psychology, 34,* 325-334.

McNeil, J. E. & Warrington, E. K. (1993). Prosopognosia: a face-specific disorder. *Quarterly Journal of Experimental Psychology, 46A,* 1-10.

Meissner, C. A. & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy and Law, 7(1),* 3-35.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience, 9,* 555-604.

Ng, W. & Lindsay, R. C. L. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology, 25(2),* 217-232.

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: a psychological and neural synthesis. *Trends in Cognitive Sciences, 6(6),* 261-266.

Perret, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells in the macaque temporal cortex. *Philosophical Transactions of the Royal Society of London, B, 335,* 23-50.

Perret, D. I., Mistlin, A. J. and Chitty, A. J. (1987). Visual neurones responsive to faces. *Trends in Neuroscience, 10,* 358-364.

Perret, D. I., Rolls, E. T. and Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research, 47,* 329-342.

Perrett, D. I., & Oram, M. W. (1993). Neurophysiology of shape processing, *Im., Vis. Comput., 11,* 317-333.

Pike, G., Kemp, R., Towell, N., & Phillips, K. (1997). Recognizing moving faces: The relative contribution of motion and perspective view information. *Visual Cognition, 4,* 409-437.

Prkachin, G. C. (2003). The effect of orientation on detection and identification of facial expressions of emotion. *British Journal of Psychology, 94,* 45-62.

Puce, A., Allison, T., Bentin, S., Gore, J.C. & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience, 18,* 2188-2199.

Rhodes, G., Brake, S., & Atkinson, A.P. (1993). What is lost in inverted faces? *Cognition, 47,* 25-57.

Rhodes, G., Tan, S., Brake, S., & Taylor, K. (1989). Expertise and configural coding in face recognition *British Journal of psychology, 80,* 313-331.

Roack, D. A., Barrett, S. E., Spence, M., Abdi, H., & O'Toole, A.J. (2003). Memory for moving faces: Psychological and Neural Perspectives on the Role of Motion in Face Recognition. *Behavioral and Cognitive Neuroscience Reviews, 2(1),* 15-46.

Rock, I. (1974). The Perception of Disoriented Figures. *Scientific American, 230,* 78-85.

Rock, I. (1988). On Thompson's inverted-face phenomenon (Research Note). *Perception, 17,* 815-817.

Rossion, B., Curran, T., & Gauthier, I. (2002). A defense of the subordinate-level expertise account for the N170 component. *Cognition, 85,* 189-196.

Rotshtein, P., Henson, R. N. A., Treves, A., Driver, J., & Donlan, R.J. (2005). Morphing Marilyn into Maggie dissociates physical identity face representations in the brain. *Nature neuroscience, 8(1),* 107-113.

Sangrioli, S. & de Schonen, S. (2004). Effect of visual experience on face processing: a developmental study of inversion and non-native effects. *Developmental Science, 7(1),* 74-87.

Sayette, M. A., Cohn, J. F., Wertz, J. M., Perrott, M. A., & Dominic. J. (2001). A psychometric evaluation of the facial action coding system for assessing spontaneous expression. *Journal of Nonverbal Behavior. 25,* 167-186.

Scapinello F. F., & Yarmey, A. D. (1970). The role of familiarity and orientation in immediate and delayed recognition of oictorial stimuli. *Psychonomic science, 21,* 329-331.

Schneider, R. U. (2008). Hurra, ich ticke nicht richtig. *NZZ-Folio, 4,* 64-73.

Schwaninger, A. & Mast, F. (2005). The face inversion effect can be explained by capacity limitations of an orientation normalization mechanism. *Japanese Psychological Research, 47(3),* 216-222.

Schwaninger, A., Lobmaier, J., & Collishaw, S. M. (2002). Component and configural information in face recognition. *Lectures Notes in Computer Science, 2525,* 643-650.

Schwaninger, A., Wallraven, C., Cunningham, D. W., & Chiller-Glaus, S. D. (2006). Processing of identity and emotion in faces: a psychophysical, psychological and computational perspective. *Progress in brain research. 156,* 321-343.

Searcy, J. H. & Bartlett, J. C. (1996). Inversion and processing of component and spacial-relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance, 22(4),* 904-915.

Sergent, J. (1984). An investigation into component and configurational processes underlying face recognition. *British journal of psychology, 75,* 221-242.

Sergent, J. (1985). Influence of task and input factors on hemispheric involvement in face processing. *Journal of Experimental Psychology: Human Perception and Performance, 11(6),* 846-61.

Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed? *Journal of applied psychology, 85(4),* 542-550.

Smith, S. M., Lindsay, R. C. L., Pryke, S., & Dysart, J. E. (2001). Postdictors of eyewitness errors: Can false identifications be diagnosed in the cross-race situation? *Psychology public policy and law, 7(1),* 153-169.

Tanaka, J. W. & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory and Cognition, 25(5),* 583-592.

Tanaka, J., Kiefer, M., & Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: evidence fron a cross-cultural study. *Cognition, 93,* B1-B9

Tanaka, J.W. & Farah, M. J. (1993). Parts and Wholes in Face Recognition. *The Quarterly Journal of Experimental Psychology, 46A(2), 225*-245.

Tanka, J.W., & Farah, M.J. (1991). Second-order relational properties and the inversion effect: testing a theory of face perception. *Perception & Psychophysics, 50(4),* 367-372.

Thompson, P. (1980). Margaret Thatcher: a new illusion. *Perception,* 9: 483-484.

Tronick, E., Als, H., & Brazelton, T. B. (1980). Monadic phases: A structural descriptive analysis of infant-mother face-to-face interaction. *Merrill-Palmer Quarterly of Behavior and Development, 26,* 3-24.

Valentine, T. & Bruce, V. (1986). Mental rotation of faces. *Memory and Cognition, 16(6),* 556-566.

Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. *British Journal of Psychology, 79(4),* 471-491.

Vernon, M. D. (1955). The function in schemata in perceiving. *Psychological review, 62,* 180-192.

Vuilleumier, P., Armony, J. L., Driver, J. and Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: an event related fMRI study. *Neuron, 30,* 829-841.

Wacholtz, E. (1996). Can we learn from the clinically significant face processing deficits, prosopagnosia and capgras delusion? *Neuropsychological Review, 6,* 203-258.

Wachsmuth, E., Oram, M. W., & Perret, D. I. (1994). Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque. *Cerebral Cortex, 4,* 509-522.

Wallraven, C., Schwaninger, A., & Bülthoff, H. H. (2005). Learning from humans: computational modeling of face recognition. *Network: Computation in Neural Systems, 16(4),* 401-418.

White, M. (2000). Parts and wholes in expression recognition. *Cognition and Emotion, 14(1),* 39-60.

Williams, M. A., Moss, S. A., & Bradshaw, J. L. (2004). A unique look at face processing: the impact of masked faces on the processing of facial features. *Cognition, 91,* 155-172.

Winston, J. S., Henson, R. N. A,. Fine-Goulden, M. R., & Dolan, R. J. (2004). fMRI-adaption reveals dissociable neural representations if identity and expression in face perception. *Journal of Neurophysiology, 92,* 1830-1839.

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). A field study of own-race bias in South Africa and England. *Psychology public policy and law, 7(1),* 119-133.

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2003). Inter-racial contact and the own-race bias for face recognition in South Africa and England. *Applied cognitive psychology, 17(3),* 365-373.

Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex, 15(8),* 1234-42

Xu, Y., Liu, J., & Kanwisher, N. (2005). The M170 is selective for faces, not for expertise. *Neuropsychology, 43,* 588-597.

Yamane, S., Kaji, S., & Kawano, K. (1988). What facial features activate face neurons in the inferotemporal cortex of the monkey? *Experimental Brain Research, 73,* 209-214.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81(1),* 141-145.

Young, A. W., Hellawell, D., Hay, D. C. (1987). Configurational information in face perception. *Perception, 16(6),* 747-759.

Yovel, G. & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current Biology, 15,* 2256-2262.

Proceedings

Essa, I. & Pentland, A. (1994). A vision system for observing and extracting facial action parameters. Proceedings of the International Conference of Computer Vision and Pattern Recognition (CVPR'94), Seattle, WA, USA, pp.76-83.

Li, J. C., Dunning, D., & Malpass, R. S. (1988). Basketball fandom and cross– racial identification among European-Americans: Another look at the contact hypothesis. Paper presented at the biennial meeting of the American Psychology-Law Society, Redondo Beach, CA.

Riegelnig, J. & Schwaninger, A. (2006). The influence of age and gender on detection performance and the criterion in x-ray screening. Proceedings of the *2nd International Conference on Research in Air Transportation, ICRAT 2006,* Belgrade, Serbia and Montenegro, June 24-28, 2006, 403-407.

Electronic Sources

Martinez, A. M. & Benavente, R. (1998). *The AR Face Database.* CVC Technical Report, 24. http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html

Fantamorph Morphing Software. http://www.fantamorph.com

Newspapers

Neue Zürcher Zeitung, Mittwoch, 9. Juli 2008

Acknowledgements

My initial thanks go to the supervisors of my thesis, Prof. Dr. Wolfgang Marx from the University of Zurich and Prof. Dr. Adrian Schwaninger from the University of Applied Sciences Northwestern Switzerland, and to the Co-referent of my thesis, Prof. Dr. Fred Mast of the University of Lausanne. I would also like to thank the Visual Cognition Research Group which accompanied me through the last four years with much valued professional advice. My special gratitude goes to Dr. Franziska Hofer who's lessons on data analysis and research design I treasure highly to this day. The Federal Police Department of Switzerland deserves my full appreciation for cooperation in this project. Also I would like to express my thanks to the class of Andreas Stirnemann, lecturer at the Zurich University of the Arts, for participation in Experiment 4.

Experiments 2, 3a-c, and 5 were financially supported by the *Stiftung Suzanne und Hans Biäsch zur Förderung der Angewandte Psychologie*; Experiments 8a and b by the *Stiftung für wissenschaftliche Forschung an der Universität Zürich*. I would like to express my gratitude to these two foundations for their generous contribution to my research.

I am greatly indebted to the four students who completed their master thesis under my supervision, Gisela Schoch, Timna Tal, Anne Bode, and Corinne Frey. They contributed tremendously to the creation of stimuli, data collection, and interpretation of the findings. Prof. Aleix M. Martinez of the Ohio State University, USA, deserves my gratitude for providing me with the priceless AR Face Database, as well as Jisien Yang, M.A., from the University of Zurich, for the large supply of Asian stimuli, Dr. Barbara Knappmeyer and Dipl. Inf. Mario Kleiner of the Max Planck Institute for Biological Cybernetics Tübingen for the stimuli of Experiment 8, and lic. phil. Markus Ruh for programming the software used in Experiment 5. Special thanks go to the countless participants of my experiments.

Finally, I would like to thank my family. My appreciation goes to Christina Glaus for motherly support, and to Dagmar Schifferli for intellectual motivation and countless valuable hours of babysitting. Most and foremost I would like to express my gratitude to my husband Michael, who shared homemaking and parental responsibilities to allow time for my research and accompanied my dissertation with never ending encouragement and inspiration. To our daughter Maya Joy, who is just learning to speak, goes all my love. Her eager, open-minded way of exploring the world provides life with a particular, deep sense of purpose.